AXES IN OBJECT-CENTERED SHAPE REPRESENTATION: INSIGHTS FROM MIRROR-IMAGE REFLECTION ERRORS

by

Thitaporn Chaisilprungraung

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

June 2020

ABSTRACT

Successful recognition and interaction with objects require the ability to perceive and represent how object parts are internally related (e.g., a teapot's handle is attached to its body, at a location opposite to the spout). According to many theorists, axes defined on the basis of object geometry provide a coordinate system for representing the locations and orientations of object parts. Understanding the function and the nature of object axes can produce key insights into theories of shape representation. This thesis examines two poorly understood problems: 1) what precisely is the mechanism by which coordinate axes accommodate shape representation and 2) what aspects of object geometry determine how axes are assigned to shapes? I address these problems in light of previous research on object orientation representation. The coordinate orientation representation (COR) theory (e.g., McCloskey, 2009) posits that the brain represents the orientation of a whole object (e.g., orientation of a pen on a floor) by encoding relations between sets of coordinate axes separately defined for the object (the pen) and for an extrinsic reference frame (the floor). The first portion of this thesis demonstrates that the COR mechanism can be adapted to explain how relationships of parts within an object are represented. The second portion uses a novel paradigm motivated by the COR theory (Chaisilprungraung et al., 2019) to investigate whether the origin of coordinate axes (i.e., the point where the axes intersect) corresponds to an individual elongated part (e.g., the handle of a hatchet) or the overall object's center (e.g., the hatchet's center of mass). The latter portion's result deepens the understanding about the mechanism for representing shapes under the adapted COR theory. These novel findings are important not only for psychological research on visual shape processing, but also for comparing the coordinate-

axis view with prominent views in computer vision research (e.g., Medial Axes theory, image recognition based on neural network theory).

Dissertation Committee

Michael McCloskey (primary advisor), Cognitive Science

Michael Bonner, Cognitive Science

Barbara Landau, Cognitive Science

Chaz Firestone, Psychological and Brain Sciences

Jonathan Flombaum, Psychological and Brain Sciences

ACKNOWLEDGEMENTS

I am grateful to a number of people without whom this work would not have been possible. I would like to thank my committee—Mike McCloskey, Mick Bonner, Barbara Landau, Chaz Firestone, and Jon Flombaum—for taking the time to read this thesis and for providing feedback. I am especially grateful to my advisor, Mike McCloskey, for his support and guidance, and his patience in teaching me to think and pushing me to be better than I could hope to be (also his patience with my English grammar). Mike is a brilliant, generous and thoughtful advisor, and it is truly a privilege to learn from him and to have him as my mentor.

I am grateful to past and present graduate students at the Cognitive Science department who created a simulating and inspiring environment where one can thrive as a researcher. My sincere thanks goes to all my undergraduate research assistants, who helped me collect data and made this thesis possible.

I would like to thank my family and friends, especially Worthy Life Baptist Church, for pouring out their love and support, and for never stopping praying for me since I began writing this thesis. Thank you Beni uni, for carrying me and my burdens all the way—I could not have completed this thesis without you. Thank you Pastor James and Donna SMN, for praying, guiding and giving me the strength to go on when I was stressed. Thank you Pastor Peter and Eunice SMN, for bringing me to your church when I was directionless in undergraduate school, and especially to Rebekah Kim JdSN, for loving me and being so mindful of my life problems as if I am your own daughter.

Most of all, I thank God for His grace, love, mercy, and faithfulness for me to complete this thesis.

In loving memory of Keow Keowrakmook, my grandfather

15 February 1933 - 28 May 2020

Whom have I in heaven but you?
And there is nothing on earth that I desire besides you
My flesh and my heart may fail
But God is the strength of my heart and my portion forever

Psalm 73:25-26

TABLE OF CONTENTS

ABSTRACT .... **ii**

ACKNOWLEDGEMENTS .... **iv**

TABLE OF CONTENTS .... **vii**

LIST OF TABLES .... **x**

LIST OF FIGURES .... **xi**

CHAPTER 1. INTRODUCTION **1**

CHAPTER 2. THEORIES OF SHAPE REPRESENTATION **5**

2.1. Structural Description Theory 5

2.2. Frame of Reference 7

2.3. Object-Based Frame of Reference: Function and Geometric Constraints 12

CHAPTER 3. FUNCTION OF SHAPE COORDINATE AXES **16**

3.1. Axes in Mathematical Coordinate Systems 16

3.2. Axes in Shape Representation 20

3.3. Medial Axes 26

3.4. A Theory of Objects' Orientation Representation 30

3.4.1. Coordinate Orientation Representation (COR) Theory 31

3.4.2. Function of Axes in The COR Theory 35

3.4.3. Orientation Recall Experiment 37

3.5. Evidence for Perpendicular Coordinate Axes in Shape Representation 39

3.6. A Potential Framework for Probing Axes' Function 42

3.6.1. Representation of Relative Part Orientation 42

3.6.2. Representation of Relative Part Location 45

CHAPTER 4. GEOMETRIC CONSTRAINTS 50

4.1. Existing Evidence for Shape Elongation 51

4.2. A Closer Look at The Notion of Elongation.. 54

4.3. Origins of Coordinate Axes 58

CHAPTER 5. INSIGHTS FROM MIRROR-IMAGE REFLECTION ERRORS....61

5.1. Goals and Motivations 61

5.2. Shape Recall Task.. 65

5.3. Experiment 1 67

5.3.1. Experiment 1 Stimuli and Methods 67

5.3.2. Experiment 1 Results 73

5.3.3. Experiment 1 Discussion 82

5.3. Experiment 2 89

5.3.1. Experiment 2 Stimuli and Methods 91

5.3.2. Experiment 2 Results 95

5.3.3. Experiment 2 Discussion 103

5.3.4. Interim Conclusion 109

5.4. Experiment 3 109

5.4.1. Experiment 3 Stimuli and Methods 111

5.4.2. Experiment 3 Results 116

5.4.3. Experiment 3 Discussion 123

5.5. Experiment 4 126

5.5.1. Experiment 4 Stimuli and Methods 127

5.5.2. Experiment 4 Results 128

5.5.3. Experiment 4 Discussion 135

CHAPTER 6. GENERAL DISCUSSION 140

6.1. Specific Mechanism for Processing Shapes?... 141

6.2. Levels of Representation.. 144

6.3. Axes for Everything? 146

6.4. Conclusion 149

APPENDIX A. Robustness of Reflection Pattern 151

A.1. Experiment 1 151

A.2. Experiment 2 152

A.3. Experiment 3 153

A.4. Experiment 4 154

APPENDIX B. Convolutional Neural Networks Analysis 155

B.1. Experiment 1 155

B.2. Experiment 2 159

APPENDIX C. Tilts of Reflection Axes 160

C.1. Experiment 3 160

C.2. Experiment 4 161

APPENDIX D. Analysis of Reflection Data Simulation 163

APPENDIX E. Manually Drawn Bounding Boxes 167

BIBILOGRAPHY 168

CURRICULUM VITA 176

LIST OF TABLES

Table D- i. Table showing coefficients that were used as bases for creating simulated responses of participants. 158

LIST OF FIGURES

Figure 2-1. (A) Feature-deleted and part-deleted stimuli in Biederman and Cooper 6

Figure 2-2. A diagram serving to illustrate how frames of reference 8

Figure 3-1. Different types of coordinate system. 16

Figure 3-2. Cylindrical and Spherical coordinate systems with only two 18

Figure 3-3. Marr & Nishihara's (1978) theory of coordinate-based shape........................22

Figure 3-4. A cow shape with medial axes 26

Figure 3-5. (A) COR diagram for representing the orientation of a knife.........................31

Figure 3-6. Representations of orientation where only the principal axis 34

Figure 3-7. Representations of orientation where only the vertical axis 35

Figure 3-8. Plot of error distribution of an orientation recall experiment 36

Figure 3-9. (A) Illustrations of how an axis of reflection is computed.............................38

Figure 3-10. (A) Illustration of the COR theory adapted for accounting for.....................41

Figure 3-11. (A) A possible coordinate-based model for representing 44

Figure 3-12. Alternative way of defining the origin of the small part's axes 46

Figure 4-1. (A) Stimuli from Quinlan and Humphreys (1993's experiment)....................48

Figure 4-2. (A) Examples of stimuli from Morikawa (1999) 51

Figure 4-3. Illustration of a hatchet showing three potential definitions 52

Figure 4-4. (A) Illustrations of how an axis of reflection is computed.............................54

Figure 4-5. Hypotheses concerning how origins of coordinate axes 56

Figure 5-1. Basic procedure of the shape recall task 62

Figure 5-2. (A) Stimulus objects in Experiment 1 64

Figure 5-3. Illustration of a trial structure...66

Figure 5-4. Illustrations of the procedures for modifying the part configuration.............67

Figure 5-5. Plots showing the tilt and location of participants' responses. 70

Figure 5-6. Histogram showing the distribution of part orientation errors........................72

Figure 5-7. Bar plot of orientation error distribution for responses..................................74

Figure 5-8. Different ways of defining the location of the small object part.....................74

Figure 5-9. Plots illustrating responses with correct locations 76

Figure 5-10. (A) Illustration of a crude form of a stimulus object 79

Figure 5-11. Predictions for location reflection errors in Experiment 2............................85

Figure 5-12. Stimuli of Experiment 2 86

Figure 5-13. Procedures for adjusting probe stimulus in Experiment 2 89

Figure 5-14. Plots illustrating responses with correct locations. 91

Figure 5-15. Plots showing participants' responses in Experiment 2. 93

Figure 5-16. Plots showing the average distances from the middle of the base part.........94

Figure 5-17. Illustration of different types of error involving the relative orientation......96

Figure 5-18. (A) Bar plot showing orientation error distribution for responses 97

Figure 5-19. COR representation underlying a 90˚-rotation error of a whole object. 100

Figure 5-20. COR representation underlying a 70˚-rotation error of an object part. 101

Figure 5-21. COR representation underlying a 110˚-rotation error of an object part......102

Figure 5-22. Stimuli for Experiment 3..107

Figure 5-23. Trial procedures of Experiment 3. 108

Figure 5-24. (A) Example showing how an axis of reflection was computed................111

Figure 5-25. All axes of reflection in Experiment 3. 112

Figure 5-26. (A). Axes of reflection for each object of Experiment 3............................114

Figure 5-27. Histogram for non-orientation reflection responses in Experiment 3.........115

Figure 5-28. Plot showing the location accuracy of participants' responses...................116

Figure 5-29. Illustrations showing an example of how a combination of different.........118

Figure 5-30. Predictions for hypotheses of Experiment 4 121

Figure 5-31. Stimulus objects for Experiment 4. 122

Figure 5-32. All axes of reflection in Experiment 4. 123

Figure 5-33. (A). Axes of reflection for each object of Experiment 4.............................124

Figure 5-34. Histogram for non-orientation reflection responses in Experiment 4.........126

Figure 5-35. Plot showing the location accuracy of participants' responses...................127

Figure 5-36. Illustrations of axes for a bottle opener when the object is unfolded..........130

Figure A- i Plots showing the distribution of orientation error in Experiment 1.............144

Figure A- ii. Plots showing the distribution of orientation error in Experiment 2...........145

Figure A- iii. Plots showing axes of reflection and the average point of.........................146

Figure A- iv. Plots showing axes of reflection and the average point of.........................147

Figure B- i Image inputs to Alex Net..148

Figure B- ii Architecture of Alex Net 149

Figure B- iii Charts displaying CNN results across conditions of..................................151

Figure B- iv Illustration of the crude forms of the stimulus objects...............................152

Figure C- i. Reflection axes of Experiment 3 displayed in a circular plot.......................153

Figure C- ii. Reflection axes of Experiment 4 displayed in a circular plot. 155

Figure D- i. Axes of reflection computed from the simulated responses 157

Figure D- ii. Plots illustrating how the tilt and the location errors were changed...........159

Figure E- i. Illustration showing bounding boxes that were manually drawn.................160

CHAPTER 1. INTRODUCTION

The ability to form internal representations of object shapes is critical for a variety of perceptual, cognitive, and motor functions. To recognize and interact with objects around us, the visual system needs access to information about the shapes of the objects. Knowledge about object shapes is also important for making sense of the visual world, and forms a basis for a host of cognitive tasks studied in the literature (e.g., perceptual judgment, mental rotation, drawing). Despite the importance of shape knowledge in visual cognition, much remains unknown about the process by which the brain analyzes shapes. Many researchers assume that, at some abstract level of visual processing, the brain represents information about shapes in terms of component parts and spatial relationships among parts (Biederman, 1987; Hoffman & Richards, 1983; Marr & Nishihara, 1978; Palmer, 1977; Singh & Hoffman, 2001). This information is assumed to be *intrinsic* in the sense that it is invariant to the circumstances in which the shape is seen. The intrinsic representation of a pen, for instance, may carry details about how the cap, the barrel, and the clip of the pen are spatially related, independent of how the pen as a whole is tilted or positioned with respect to the observer or anything else in the environment. To put it another way, the brain represents shapes in an object-intrinsic frame of reference (i.e., an object-centered representations).

Critical to the idea of a frame of reference is the notion of object axes. Many researchers assume that the intrinsic relationships among object parts (or features) are defined with respect to a single most important axis of the object, the so-called “principal axis” (Jolicoeur & Humphrey, 1998; Leek & Johnston, 2006; Quinlan, 1991; Quinlan & Humphreys, 1993; Tarr & Pinker, 1990). The cap, the barrel, and the clip of the pen may be defined in terms of where they are relative to the principal axis of the pen, which is defined to be parallel with the pen's elongation. In short, object axes are thought to provide some sort of coordinate system within which internal part relations can be defined.

Many studies on shape cognition assumed that insights about shape representation can be obtained by examining the nature of object axes (Boutsen & Marendaz, 2001; Herbert et al., 1994; Large et al., 2003; Ling & Sanocki, 1995; Quinlan & Humphreys, 1993; Sekuler, 1996; Sekuler & Swimmer, 2000). These studies aim to discover what geometric properties of a shape constrain how the principal axis is defined for an object. Based on existing findings, the principal axis may be defined in parallel with shape elongation or shape symmetry (Herbert et al., 1994; Ling & Sanocki, 1995; Morikawa, 1999; Quinlan & Humphreys, 1993; Sekuler, 1996; Sekuler & Swimmer, 2000; Smith et al., 2014).

The aim of this thesis is to expand upon the current understanding of object axes. Many researchers paid attention to the geometric properties that constrain how the axes are defined (henceforth: "geometric constraints"). However, another question deserves equal consideration. One may ask what precisely is the role and the mechanism by which the axes accommodate the representation of internal part relationships. I refer to this question as the question about *function.* As will be later demonstrated, insights about the axes' function are equally important for understanding the nature of object shape representation.

This thesis is organized into six chapters. Following the present introductory chapter, Chapter 2 provides theoretical background necessary for understanding the roles of object axes in theories of shape representation. In chapter 3, I expound questions concerning the axes' function. I suggest that the main function of axes is to provide bases for defining spatial directions, which are important for establishing relationships among parts within an object. Based on this understanding of axis function, I claim that object shapes must be defined with respect to multiple coordinate axes. I consider a theory of orientation representation, which posits that the overall orientation of an object (e.g., orientation of a teapot in a room) is mentally represented by relating the object-intrinsic frame of reference (e.g., the teapot's frame) to a frame of reference extrinsic to the object (e.g., the room's frame). Critically this theory--the Coordinate Orientation Representation (COR) theory--characterizes frames of reference in terms of perpendicular coordinate axes (McCloskey et al., 2006). I consider ways in which the COR theory may be adapted to account for representing relationships of parts within an object (e.g., the orientation and the location of a teapot's spout relative to the teapot). The ideas developed in this chapter are later tested in two experiments of the thesis.

In chapter 4, I reconsider questions about how object axes are defined. I point out that despite a large body of research on geometric constraints, some aspects of the problem remain unresolved. Previous research demonstrated that the tilts of object axes are defined based on shape elongation. The principal axis of a pen, for instance, is defined so that it is parallel with the pen's elongation (Ling & Sanocki, 1995; Morikawa, 1999; Sekuler & Swimmer, 2000; Smith et al., 2014). For objects with more than one part (e.g., a hatchet with a blade and a handle), a study suggested that the basis for defining axes' tilts corresponds to the object's longest part (i.e., the handle) rather than to the object's overall shape (Chaisilprungraung et al., 2019) Under the assumption where perpendicular axes are involved in shape representation, however, it is also important to understand how the origin of the coordinate axes (i.e., the point where the perpendicular axes meet) is defined. Understanding how axes' origins are defined can be critical to theories of axis-based shape representation. Currently however, there is very limited discussion on this issue.

Chapter 5 discusses four experiments bearing evidence for functions and geometric constraints of shape axes. Results from the first two experiments suggest that the COR mechanism for representing whole objects can indeed be adapted to explain how relations of object parts are represented, thus demonstrating the importance of coordinate axes in shape representation. In the latter two experiments, I demonstrate that the basis for defining the origin of coordinate axes corresponds to the overall object rather than to a single object part, and interpret this result in light of possible roles of coordinate axes. Finally Chapter 6 relates the current findings to broader issues concerning the specificity of the COR-based mechanism, levels of shape processing, and relevance of axes to the processing of non-shapes.

CHAPTER 2. THEORIES OF SHAPE REPRESENTATION

2.1. Structural Description Theory

A prominent theory of shape representation holds that the brain stores shape information in terms of descriptions of how the structural components are internally related. The structural description theory was posited as an attempt to understand how the visual system recognizes shapes, and many of the theory's details deal with the process of recognition. Vision, according to some prominent theorists such as Marr (Marr, 1982; Marr & Nishihara, 1978), is a process whereby arrays of light intensity are transformed into symbolic representations. These perceptual representations store details about how object parts are internally related. In recognition, the perceptual representations must somehow be matched to the corresponding representations of known objects stored in memory.

Arguments for the structural description theory were largely obtained from studies on visual object recognition. Biederman and Cooper (1991), for instance, claimed that recognition involves extraction of object parts. They created line drawings of objects similar to those shown in Figure 2-1A, where the objects' contours had been deleted in one of two ways. In the 'feature-deletion' condition (Figure 2-1A, top row), every other feature of the object (i.e., edge, vertex) was deleted from the contour, resulting in the removal of half of each object part. In the 'part-deletion' condition (Figure 2-1A, bottom row), every other object part was deleted, resulting in the removal of half of the object parts. The researchers were interested in the priming effect of these stimuli on subsequent recognition. Participants first viewed these contour-deleted line drawings during the first block of the experiment. In the second block, they saw contour-deleted images that were either exactly identical to the ones in the first block, or complementary to them (in Figure 2-1A, complementary drawings were shown side by side in the same row). Participants were instructed to name objects in the second block as quickly as possible.

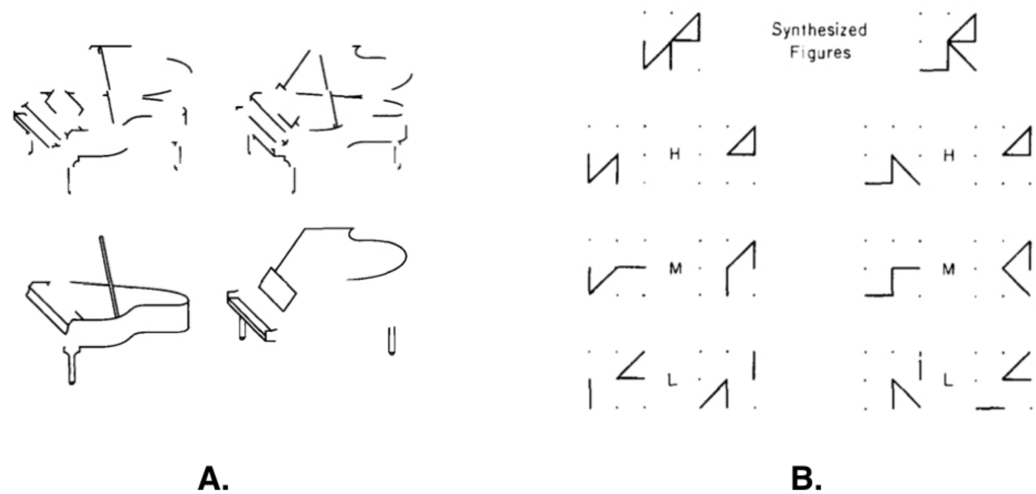

Figure 2-1. *(A) Feature-deleted and part-deleted stimuli in Biederman and Cooper (1991)'s study. (B) Examples of stimuli from Palmer (1977)'s study.*

For stimuli in the 'feature-deleted' condition, Biederman and Cooper observed a high level of priming (i.e., higher accuracy, lower RT), regardless of whether the subsequent contour-deleted drawings were identical or complementary. For the 'part-deleted' condition, however, the priming effect was high only when the subsequent stimuli were identical. The fact that minimal priming was observed for complementary stimuli in the 'part-deleted' condition, but not the 'feature-deleted' condition, led the researchers to suggest that the recovery of object parts, but not the recovery of specific object features, was important for priming in object recognition. More broadly, Biederman and Cooper suggested that recognition of object shapes involved the extraction of object parts.

Besides evidence for the perceptual representations of object parts, support for the structural description theory was also derived from arguments that object parts were represented in a hierarchical fashion. The notion that object parts are hierarchically organized implies the idea of selective organization—or the idea that certain elements in a visual stimulus are more likely to behave as a group, or as a structural unit, than others. Palmer (1977) demonstrated the hierarchical nature of shape representation by asking observers to rate the 'part' status of stimulus figures as those shown in Figure 2-1B. These stimulus figures were made up of three line segments formed by connecting the dots in the 3x3 grid in different ways. Palmer demonstrated that line segments grouped in certain ways were more likely to be rated as having a 'high' part status (i.e., see the top row in Figure 2-1B), whereas others as 'medium' or 'low' status (i.e., see the middle and the bottom row in the figure). The ratings of part status were highly consistent across participants. Moreover, the ratings successfully accounted for data of later experiments in the study, where a separate group of participants engaged in some other perceptual judgment tasks that involved the same set of stimuli (e.g., deciding whether two stimulus figures could be synthesized into a unitary shape). The evidence suggested that, at least for the type of stimuli used in Palmer's study, certain components of a shape are more readily identified as forming a unit than others.

2.2. Frame of Reference

A critical assumption of the structural description theory is the notion of a spatial frame of reference. Researchers discussing structural relations often use terms such as "right of", "west of", "above", "adjacent to", "behind", or "tilted clockwise" to refer to how part relationships are encoded (Barlow, 1972; Connell, 1985; Foster, 1982; Quinlan, 1991; Sutherland, 1968; Winston, 1970). These terms are meaningful only if they are defined in some frame of reference (Hinton, 1981). To appreciate the concept of reference frame, it is helpful to consider a simple problem of how to define an object's location in the environment.

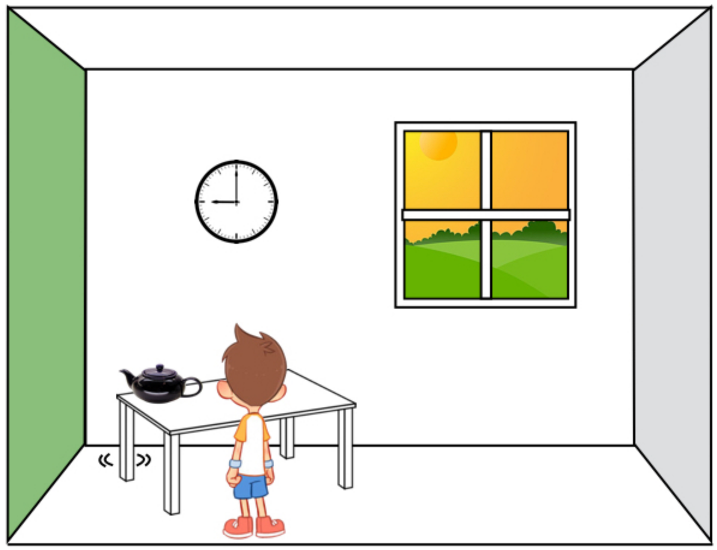

Figure 2-2. *A diagram serving to illustrate how frames of reference are relevant for defining spatial locations.*

Consider how the location of the teapot in Figure 2-2 may be defined. Any term that describes the teapot's location (e.g., "left of", "on top of", "near to", "inside") must be defined based on something external to the teapot. For instance, we may say that the teapot is "on top of" the table, and "close to" the corner of the tabletop with a wobbling leg. In this description, the teapot's location is defined relative to the features of the table. That is, the location is defined in *a table-based frame of reference*. Another possible description of location may be based on the room's layout: The teapot is "inside" the room, "nearest to" the green wall. In this case, a *room-based frame of reference* is involved in defining the location of the teapot. A number of objects, living things, or features of the environment (e.g., direction of gravity, direction of sunrise) can serve as possible frames of reference for defining locations in the real-world environment. Another possible basis for defining locations is the observers themselves. The location of the teapot in Figure 2-2 may be defined as being directly in front of the observer (i.e., the boy), with the center of the teapot aligned with the midline of the observer's body. In this thesis, the latter type of frame of reference will be referred to as an "observer-centered reference frame", and all other extrinsic frames as "environment-centered reference frames."

In the example above, frames of reference provide bases for defining locations of objects in an environment. In the structural description theory, frames of reference are important in so far as they allow internal relationships among object parts to be defined. Descriptions about shape structure must include information about how object parts are positioned relative to one another or relative to the entire object (e.g., the teapot's spout is adjacent to the pot's body). In addition, information about the relative tilt or relative orientation must somehow be specified (e.g., the spout is tilted at an angle relative to the overall object).

As in the example of the teapot's location, structural relations within an object may be defined with respect to different frames of reference. A common type of observer-centered reference frame that is thought to be relevant for describing relationships among object parts is one that is defined by where the observers' eyes are fixated. In a *retina-based frame of reference,* descriptions of part relations are typically defined based on some abstract horizontal and vertical axes centered on the point of eye fixation (Epstein & Lovitts, 1985; Sutherland, 1968). If the observer (i.e., the boy) in Figure 2-2 fixes his gaze at the center of the teapot, various parts of the pot may be defined with respect to the axes of the retina. The round knob, for instance, may be defined as being the farthest from the location of fixation along the vertical axis of the retina. The teapot's spout may be defined as having the furthest distance along the horizontal axis, with the spout's elongated structure tilted roughly 45˚ with respect to either of the axes. The coding of shape structure in a retina-based frame of reference changes with fixation: If the observer in Figure 2-2 tilted his head to the left 90˚, the round knob of the teapot must now be defined as being the farthest along the horizontal axis of the retina. The teapot's spout and the handle, on the other hand, are defined as being on at the opposite ends of the vertical axis.

Relations of object parts may also be encoded with respect to some features of the environment. A common type of environment-centered frame is one that is defined based on the direction of gravity (Attneave & Olson, 1967; Rock, 1973, 1983). In a *gravity-based frame of reference*, part relations are defined relative to the axis perpendicular to the flat surface of the earth. Although often omitted in the discussion of this literature, a gravity-based frame must also be, to a certain extent, observer-centered. Rock (1973) for example discussed a gravity-based frame in terms of perpendicular axes, which consisted of a vertical axis parallel with the gravitational direction, and two other axes that clearly must be defined by something other than gravity. That is, Rock assumed that there was also a left-right axis and a front-back axis, which might be defined based on the left vs. right and the front vs. back sides of the observer's body (Rock, 1973, 1983). In a gravity-based frame, the lid of the teapot may be defined as being "topmost" relative to all other parts of the object, and the tip of the spout as the "leftmost". Unlike the retina-based coding, the gravity-based coding does not change with fixation: Regardless of how the observer in Figure 2-2 tilts his head or where he directs his gaze, the lid remains the "topmost" part of the object and the spout as the "leftmost" part. However, the gravity-based coding changes with the overall tilt of the object in the environment: If the teapot is rotated 90˚ counterclockwise, such that the tip of the spout points to the ground instead of to the left; then the handle, and not the lid, must now be defined as being "topmost". The lid and the spout are now defined as "leftmost" and "bottommost", respectively.

Finally, structural relations may be represented in a frame of reference based on the object itself. In an *object-centered frame of reference*, part relations are defined along some directions intrinsic to the object. Researchers often characterized these directions in terms of axes. The relative location of the spout and the handle of the teapot, for instance, may be defined as being on the opposite side of the teapot along the axis (or the direction) parallel with the elongated structure of the teapot.[1] The orientation of the spout can also be defined as having a tilt approximately 45˚ relative to the axis of the teapot's elongation.

There are well-known advantages of structural descriptions that are defined in object-centered frames of reference. Part relations that are defined intrinsically to an object do not change with changes in the observer's perspective or point of fixation, or with the tilt variations of the object. If the observer in Figure 2-2 changes where he looks at the teapot, or if the teapot is rotated in some ways with respect to the environment, the coding of part relations remains the same. If shape representations are stored in a format where only intrinsic part relations are preserved, and not circumstantial details associated

[1] If this axis is drawn as a line over the teapot in Figure 2-2, it will correspond to a horizontal line.

with how an object is viewed (e.g., observer's viewpoint), then only one representation corresponding to the canonical structure of an object is needed. If on the contrary, shape representations are defined relative to the observers' viewpoints, then numerous representations corresponding to different viewpoints may be required, rendering visual recognition a challenging problem (Biederman, 1987; Marr & Nishihara, 1978).[2]

2.3. Object-Based Frame of Reference: Function and Geometric Constraints

The previous section provided a broad overview of different types of reference frames in shape representation. As a working assumption of this thesis, I assume that at some high level of shape processing, the brain stores representations of object shapes in an object-centered frame of reference. The primary focus of the thesis is to investigate the nature of these object-centered shape representations.

As the previous section suggested, a number of theorists characterized object-centered frames of reference in terms of axes. In studies of object axes, it is assumed that there exists a single most important axis, the so-called "principal axis", which is defined on the basis of some feature(s) of the object (Jolicoeur & Humphrey, 1998; Leek & Johnston, 2006; Quinlan, 1991; Quinlan & Humphreys, 1993; Tarr & Pinker, 1990). The principal axis of a pen, for instance, may be defined so that the axis is parallel with the elongated structure of the pen. Researchers assumed that the object principal axes provide some sort of coordinate system for representing objects' structures. Marr (Marr,

[2] This is not to say that there is no disadvantage for theories that assume object-centered shape representations. Objects seen from a certain perspective can have certain parts occluded behind other parts or structures of the object. For these occluded views, it still remains to be solved how shape representations are computed from object images.

1982; Marr & Nishihara, 1978), for instance, suggested that the relative locations and orientations of object parts may be encoded in coordinate systems (i.e., the cylindrical and the spherical coordinate systems) that are built around principal axes (details of Marr's proposal will be discussed later). By understanding how the principal axis was defined for a given object, researchers assumed that insights about object shape representation could be obtained (Boutsen & Marendaz, 2001; Herbert et al., 1994; Large et al., 2003; Ling & Sanocki, 1995; Quinlan & Humphreys, 1993; Sekuler, 1996; Sekuler & Swimmer, 2000).

Although much attention in research is given to the object principal axis, there are reasons to think that other axes are also necessary in shape representation. To represent the location of a given point in a mathematical coordinate system, for instance, the location is typically defined relative to multiple perpendicular axes (i.e., axes typically denoted as X, Y, Z in the Cartesian coordinate system). Researchers on shape representation do not necessarily disagree with this notion of multiple axes. However, in this literature there is little open discussion about what roles axes play in representing shape structures, and therefore it is not always clear why multiple axes are necessary in shape representation.

In the first part of this thesis, I clarify the roles of axes in theories of shape representation. I first consider common types of mathematical coordinate systems (i.e., Cartesian, cylindrical, spherical coordinate systems), and examine the roles that axes play in accommodating spatial representation in each system. As will be argued, the primary function of axes is to provide bases for defining *directions* (§ 3.1. Axes in Mathematical Coordinate Systems). If locations are represented in 3D, the presence of three coordinate axes is required. If locations are defined in 2D, two axes are required. Based on this understanding of coordinate axes, an argument is developed that object shapes must be represented with respect to multiple axes, since shapes can only be defined in 2D or 3D (§ 3.2. Axes in Shape Representation). In § 3.3, I discuss a view on axis-based shape representation (i.e., medial-axis view), which embraces a set of theoretical assumptions rather different from the coordinate-based view I have been discussing. This section also demonstrates that insights about axes' functions can be very useful for comparing different accounts of shape representation. Finally, in § 3.4 to 3.6 I discuss a theory on objects' orientation representation. The COR (Coordinate Orientation Representation) theory posits how the overall orientation of an object in the environment (e.g., orientation of a teapot in a room) may be defined in coordinate systems with perpendicular axes (McCloskey, 2009; McCloskey et al., 2006). I consider how the COR theory can be adapted to account for the representation of part relations within an object. The ideas developed in these final sections provide bases for the first two experiments proposed in chapter 5.

In the second part of the thesis, I examine what geometric features (i.e., geometric constraints) of shapes are important for determining how object axes are defined. Previous research suggested that the object principal axes were defined so that they were parallel with shape elongation or shape symmetry (Herbert et al., 1994; Ling & Sanocki, 1995; Morikawa, 1999; Quinlan & Humphreys, 1993; Sekuler, 1996; Sekuler & Swimmer, 2000; Smith et al., 2014) (§ 4.1. Existing Evidence for Shape Elongation). To some extent, however, the identification of 'shape elongation' depends on what shapes are being considered. Given a simple straight object such as a pen, shape elongation is easily defined as the axis parallel with the long dimension of the pen. Given a more complex object, say, a hatchet (consisting of a long handle and a blade attached at one end of the handle), the principal axis might be defined so that it is parallel with the object's longest part (i.e., the handle), or it might be defined in some way which takes into account the entire object's structure (i.e., along a direction that spans both the handle and the blade). In my previous study I obtained evidence supporting the elongated-part hypothesis (Chaisilprungraung et al., 2019) (4.2. A Closer Look at The Notion of Elongation). In this study there was also evidence that object structures were represented with respect to perpendicular coordinate axes. However, while the finding might reveal interesting insights about how the *tilts* of the axes were defined, it was not known what geometric features of the objects determine how the *origin of the coordinate axes* was defined. The point where the perpendicular axes intersect, for example, may coincide with the center of the elongated part, or it might coincide with the center of the overall object. In the two latter experiments of this thesis, I propose a method for testing these hypotheses.

CHAPTER 3. FUNCTION OF SHAPE COORDINATE AXES

3.1. Axes in Mathematical Coordinate Systems

A variety of mathematical coordinate systems exist for representing spatial locations. In mathematics, the term 'coordinate system' refers to a system where a given point is determined by a set of numbers. These numbers correspond to some types of distance (e.g., angular distance, linear distance) that together define the spatial location of the point. Importantly, the numbers (henceforth: coordinate values) must be defined relative to some axes.

Consider, for instance, a Cartesian coordinate system. In a Cartesian coordinate system, each coordinate value in an ordered pair (x, y) or a triple (x, y, z) corresponds to a linear distance from a reference point (origin) on an axis (e.g., Axis X, Y, Z[3] in Figure 3-1). In the triple $(-4,5,3)$, for example, the coordinate values correspond to the linear distances of four units on the X-axis, five units on the Y-axis, and three units on the Z-axis. The signs (+ or -) also signify directions relative to the center of a coordinate system (i.e., the point represented by the triple $(0,0,0)$, or the point where all the axes meet). The four-unit distance is defined in the direction given by the – pole of the X-axis, and the five- and the three-units in the directions given by the + poles of the Y- and the Z-axis, respectively (Figure 3-1). In a Cartesian coordinate system, axes provide bases for defining *spatial directions.* The directions, in turn, are important for establishing the location of a given point.

[3] In the thesis, capital letters are used for notating axes, and lower-case letters for coordinate values.

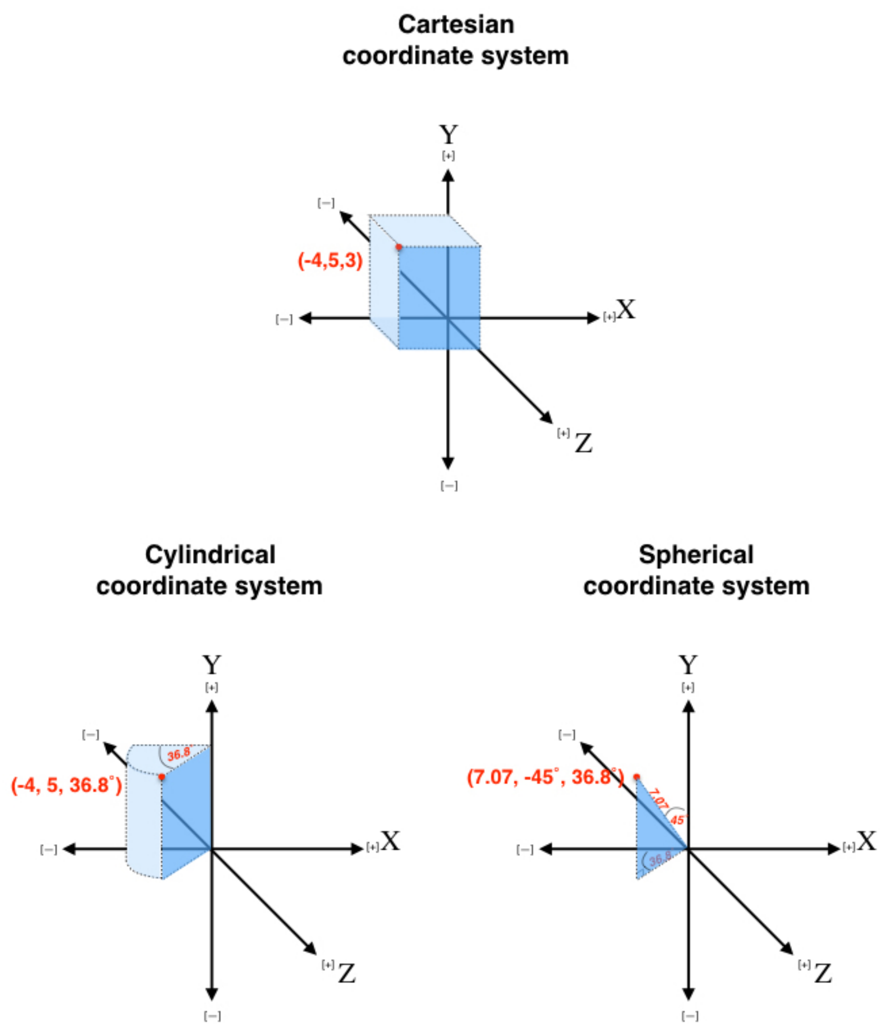

Figure 3-1. *Different types of coordinate system.*

In other types of coordinate systems, axes also provide bases for defining spatial directions. Figure 3-1 shows how the location defined by the Cartesian coordinate $(-4,5,3)$ may be represented in a cylindrical and a spherical coordinate system (these locations are depicted as red dots in the figure). In the cylindrical coordinate system, the first two values in the triple $(-4,5,36.8^{\circ})$ correspond to linear distances on the X-axis and the Y-axis (i.e., four units on X-axis and five units on Y-axis). The last coordinate value (36.8°) corresponds to the angle of rotation, which is formed by sweeping the rectangular area (i.e., the 2D area formed by the x- and the y-coordinates; highlighted in blue in the figure) about the Y-axis, and in the direction given by the [+] pole of the Z-axis. In the spherical coordinate system, the first coordinate value (7.07) corresponds to the radial distance, or the absolute distance to the center of the coordinate system. The second value (-45°) corresponds to the angle with which the radial axis is inclined against the Y-axis. The direction of the inclination is specified by the [-] pole of the X-axis. The third value (36.8°) corresponds to the angle with which the triangular area (i.e., the area formed by the radial axis and its orthogonal projection; highlighted in blue) is rotated about the Y-axis. The direction of rotation is given by the [+] pole of the Z-axis.

Perhaps the best way to appreciate the function of axes in a cylindrical and a spherical coordinate system is by removing one of the axes and considering the consequences. In Figure 3-2, the locations given by the red dots are defined relative to two coordinate axes (i.e., the X-axis and the Y-axis). With the Z-axis missing, there is no basis for specifying the direction of the angle of rotation. That is, it is not possible to distinguish whether the rectangular and the triangular areas (i.e., in the cases of the cylindrical and the spherical coordinate system, respectively) are rotated into or out of the picture plane (i.e., the plane formed by the X- and the Y-axis). As a natural consequence, a coordinate triple (x, y, z) may correspond to *either* one of two locations that are reflected about the xy plane. Axes in these coordinate systems serve as bases for defining spatial directions. To define *unique* locations in 3D, the presence of three axes is required.

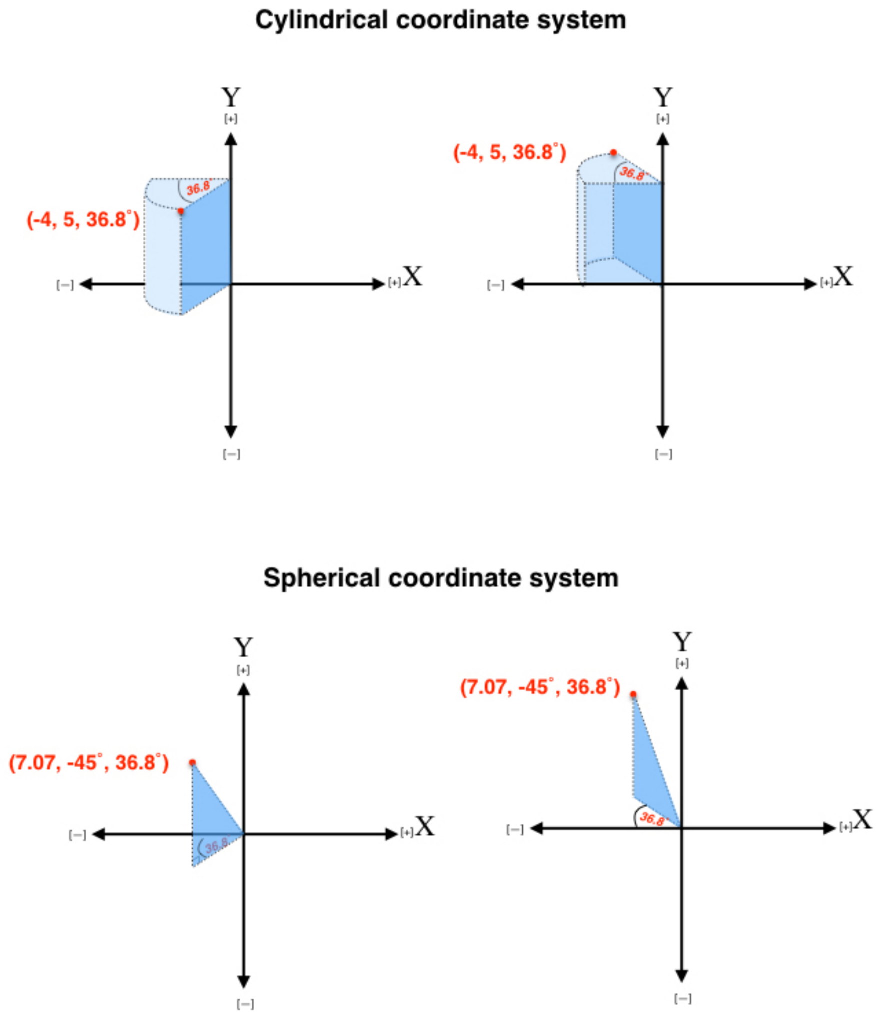

Figure 3-2. *Cylindrical and Spherical coordinate systems with only two coordinate axes*

Other types of mathematical coordinate systems exist besides the three-aforementioned systems. Some systems define locations in terms of linear distances only (e.g., Barycentric coordinate system, Trilinear coordinate system), whereas in others angular distances or more complex types of distance are involved (e.g., Generalized coordinate system, Curvilinear coordinate system, Parabolic coordinate system). Regardless of the particular types of coordinate system, the basic function of the axes is the same: Axes provide bases for defining spatial direction, which is required for establishing any type of distance in the Euclidean space. The form of coordinate axes does not depend on the type of coordinate system, but on whether locations are defined in 2D space or 3D space. If locations are defined in the two-dimensional space, two axes are required; if they are defined in three dimensions, three axes are required.[4] The diagrams in Figure 3-1 depicted the coordinate systems with perpendicular axes. It should be noted that perpendicularity is not necessarily a requirement of a coordinate system. In the curvilinear coordinate system, for instance, locations are defined relative to skew axes (i.e., axes that are not parallel but also not perpendicular with one another). The number of skew axes required in a coordinate system is also dependent on the number of dimensions involved in representing locations (i.e., two skew axes are required for 2D locations and three skew axes for 3D locations).

3.2. Axes in Shape Representation

The previous section discussed the function of coordinate axes in representing the location of an individual point in space. In shape representation, axes play critical roles in accommodating the encoding of relations within a shape. Relations such as the relative position, orientation, or length of a part within an object must be encoded along multiple independent directions. Coordinate axes fulfil this requirement by providing bases for defining the directions.

As a concrete illustration of the roles of axes, consider how the shape of a teapot may be represented. One way to relate different parts of the teapot (e.g., spout, lid,

---

[4]An exception to this is the curvilinear coordinate system, where locations are represented on the basis of curved axes or skew axes. In the curvilinear system with curved axes, locations in 2D may be represented with respect to a single curved axis (and locations in 3D with respect to two curved axes). For the purpose of this thesis, this type of coordinate system will not be considered.

handle) is to specify their relations along multiple directions in the object's intrinsic frame of reference. A possible reference frame of the teapot may include an axis parallel with the teapot's overall elongation (i.e., the axis that extends across the spout, the round body, and the handle of the teapot), an axis parallel with a line that can be drawn from the lid to the base (i.e., the axis orthogonal with the teapot's flat bottom), and an axis perpendicular to the two former axes. Relations of the teapot's parts are encoded with respect to these axes. With respect to the axis of elongation, the location of the spout may be encoded as being opposite the handle, and adjacent to the round body of the pot. With respect to the lid-base axis, the spout may be encoded as being approximately at the same location as the handle and the round body, and somewhere between the lid and the flat bottom of the teapot. It is important to note that different types of information are available along different axes. Therefore to obtain the spatial descriptions of an object in three dimensions, three axes are required.

Given the discussions thus far, the role of axes in shape representation may seem obvious. Nonetheless, among psychological studies where the notion of object-centered frame of reference is assumed, there is little discussion about how exactly coordinate axes accommodate the representation of shape structures. In these studies, researchers assume that there is a single most important axis for an object's frame of reference, the so-called "principal axis", and theoretical and empirical discussions are conducted following this notion (Large et al., 2003; Leek & Johnston, 2006; Ling & Sanocki, 1995; McMullen & Farah, 1991; Quinlan & Humphreys, 1993; Rock, 1973; Sekuler, 1996; Sekuler & Swimmer, 2000; Tarr & Pinker, 1990; Wilson & Farah, 2003). In an experiment by Quinlan and Humphreys (1990), for instance, participants were asked to draw a line they felt went naturally with abstract, two-dimensional, line-drawn geometric shapes (see Figure 4-1 in chapter 4 for the complete set of stimuli). The underlying assumption of the study was that the brain defines a single most important axis for an object-centered frame of reference. By asking participants to select a line they felt went most naturally with a shape, it was assumed that the representation of this axis (and thus the representation of the object's frame of reference) could be revealed. In Chapter 4 of this thesis, I discuss this study and its theoretical implications in more detail. The strong emphasis on the principal axes can perhaps be traced back to Marr, whose ideas inspired later research. Marr's critical idea was that the visual system defined a principal axis for each object part or group of parts (e.g., the finger, the hand, the forearm, the leg of the(Marr & Nishihara, 1978) human body; see Figure 3-3A). In Marr's proposal, the principal axis has two functions. The first function is to provide a basis for defining a structural unit of a shape (i.e., the so-called "generalized cone"). Mathematically, a generalized cone is a piece of 3D volume created by sweeping a cross section of constant shape but varying size along a principal axis. The second function is to provide a *coordinate system*: In Marr's proposal, the axis in a generalized cone can also act as a coordinate system for relating parts in an object (Figure 3-3B&C). However, as I argued earlier, a coordinate system for relating components within a 3D shape must include three axes. I will return later to analyze Marr's claim about this latter function of the principal axis.

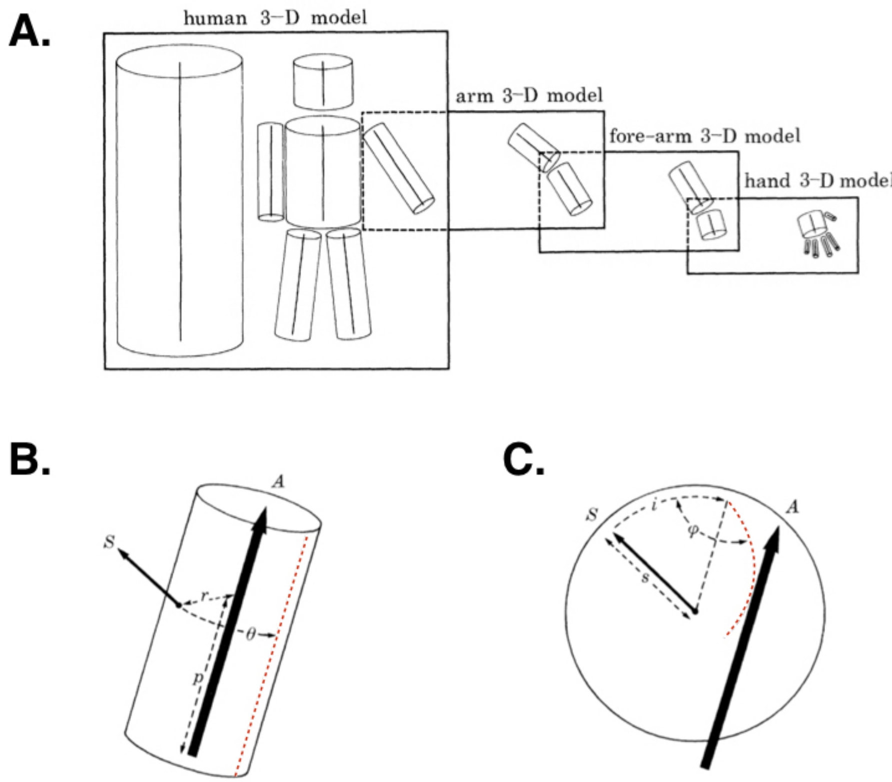

Figure 3-3. *Marr & Nishihara's (1978) theory of coordinate-based shape representations. Illustrations show (A) how relationships among parts of a shape are encoded by defining a principal axis for each part of the shape, and (B&C) how relations between two principal axes may be encoded.*

Marr also assumes that relationships among parts are represented in a hierarchical fashion (Marr & Nishihara, 1978; see also: Hoffman & Richards, 1983; Palmer, 1977). Parts that were more local were represented relative to a more global part (e.g., the fingers were represented with respect to the hand), and parts that were global at one level of hierarchy might become local at another (e.g., the hand was global relative to the fingers, but was local relative to the forearm). To represent the global-local relationships of object parts, the visual system relates the principal axis of the local part (e.g., Axis S; Figure 3-3B&C) to the principal axis of the global part (e.g., Axis A in the figures). Here the principal axis of the global object part acts as (part of) a *coordinate system* for defining spatial relations for the part(s) at the immediately lower level in the hierarchy. The relation of *location* (e.g., the location of Axis S relative to Axis A; Figure 3-3B) is defined in a cylindrical coordinate system. The relation of *orientation* (the orientation of Axis S relative to Axis A; Figure 3-3C) is defined in a spherical coordinate system.

In Figure 3-3B&C, the location and the orientation of the local object part are defined relative to one axis in a coordinate system. Although Marr did not specify whether any other axis besides the principal axis was required in a coordinate system, there was some indication that something else besides the principal axis must exist in order to have a complete representation of part relations. Particularly, some components of coordinate-based shape representation must be defined based on directions other than the direction aligned with the principal axis. In the cylindrical coordinate system (Figure 3-3B), the relative location of the smaller part's principal axis (i.e., Axis S) is defined by specifying where an axis' end point (i.e., the end point of Axis S) is relative to Axis A. The parameter p and r specify the linear distances of the end point from one extreme end of Axis A.[5] The parameter θ specifies the amount of rotation of the end point about Axis A, relative to some unspecified axis (i.e., the red dash axis[6] displayed on the surface of the cylinder). In this diagram it is not possible to determine the *unique* location of the end point about Axis A without knowing *which direction* the rotation (denoted with the parameter θ) takes place about Axis A. This problem is analogous to the problem I discussed earlier in the previous section, where the location of the red dot in the

[5] The linear distance parallel with Axis A is specified by the parameter p. The distance orthogonal with Axis A is specified by the parameter r.

[6] This axis (as well as the axis in the spherical coordinate system; Figure 3-3C) was colored in red in the current thesis for the purpose of illustration.

cylindrical coordinate system was defined only relative to the X-axis and the Y-axis (see Figure 3-2). Lacking the Z-axis (or some prior means for establishing a direction parallel with the Z-axis' direction), it was not clear how to define which way the rotation proceeded about the Y-Axis, and thus it was not possible to determine a unique location about the Y-axis. Another problem with Marr's diagram is that it is not clear what might serve as the starting point for defining the rotation (i.e., θ in Figure 3-3B) about Axis A. In the figure the rotation (θ) is shown as having a starting point corresponding to a location on a red dashed line, but it is unclear how this reference location (or the line itself) may be a priori defined with respect to the principal axis. To establish *any* unique location about a principal axis, something else other than the principal axis must be present to give basis for defining spatial directions. The same analyses apply to the spherical coordinate system in Figure 3-3C. In this figure the relative *orientation* of Axis S is defined by relating the other end point of Axis S (i.e., the end point with an arrowhead) to Axis A. The parameter ϕ specifies the amount of rotation of the end point about Axis A. Here too, the rotation must be defined based on some directions and with respect to a starting point that must be established independently of Axis A.

A coordinate system with three perpendicular axes (e.g., the axes depicted in the diagrams in Figure 3-1) naturally fulfils these requirements. Each axis in the perpendicular set serves as an independent basis for a direction, making it possible for a unique location in 3D space to be defined. It should be noted that in bringing up these discussions I do not intend to suggest that Marr denied the possibility of axes or something else other than the principal axis in an axis-based 3D model of shape representation. However, if the notion of multiple axes was assumed in his proposal, he did not make it clear (Marr, 1982; Marr & Nishihara, 1978). The same may also be said for other psychological studies that assume the notion of object-centered reference frame (Large et al., 2003; Leek & Johnston, 2006; Ling & Sanocki, 1995; McMullen & Farah, 1991; Quinlan & Humphreys, 1993; Rock, 1973; Sekuler, 1996; Sekuler & Swimmer, 2000; Tarr & Pinker, 1990; Wilson & Farah, 2003).

3.3. Medial Axes

The previous section demonstrated how insights about the axes' *function* might be useful for clarifying the nature of object-centered shape representations. Another potential significance of the insights is that they provide a solid framework for comparing the coordinate-axis account with another account of shape axes that embraces a different set of theoretical assumptions. According to the Medial-Axis Theory, the visual system computes *medial axes* from the contour of a shape (Ayzenberg et al., 2019; Blum & Nagel, 1978; Firestone & Scholl, 2014; Harrison & Feldman, 2009; Kovács et al., 1998; Lowet et al., 2018; Palmer & Guidi, 2011; Van Tonder et al., 2002; Wilder et al., 2011). A medial-axis representation consists of multiple connected line segments of varying length, formed by points that are equidistant from the contours of a shape (see Figure 3-4 for examples). Various algorithms have been developed for computing medial axes, each of which yields a somewhat different version of axes for a given shape (Blum & Nagel, 1978; Bradshaw & O'Sullivan, 2004; Cornea et al., 2007; Dey & Sun, 2006; Feldman & Singh, 2006; Lee et al., 1994; Siddiqi & Pizer, 2008). An intuitive grasp of the medial-axis concept can perhaps best be gained from Blum's grassfire algorithm (Blum & Nagel, 1978). Imagine simultaneously setting fire to all boundary points on a shape. Assuming that the fire spreads at a constant rate, the medial axes are formed by all points where the fire converges.

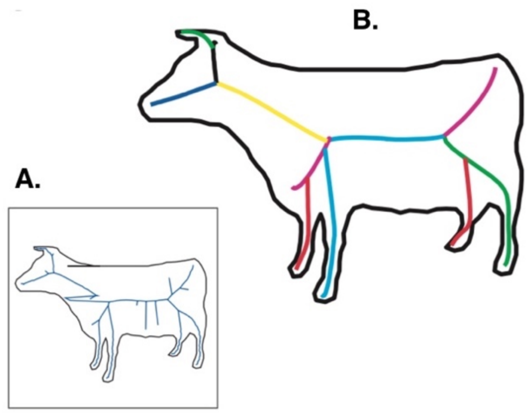

Figure 3-4. *A cow shape with medial axes computed by (A) the original grassfire algorithm proposed by Blum (Blum & Nagel, 1978), and (B) a more sophisticated algorithm developed by a later study (Feldman & Singh, 2006). Figures from Feldman & Singh (2006).*

In the medial-axis viewpoint, theorists assume that an object shapes have some interior structures, what are often called "shape skeletons", that can be modeled by axes computed from the shape's contour. The notion of "skeleton" is based on the idea of *local symmetry*—or the idea that the axis of symmetry can be somehow identified from a local part within a shape. The aim of many medial-axis AI researchers is to design an algorithm producing axes that closely approximate this notion of "shape skeletons". The axes computed after Blum's early grassfire definition[7] were often criticized on the ground that they did not always correspond to shape skeleton in an intuitive way (e.g., see the axes of the cow in Figure 3-4A). Later medial-axis researchers, recognizing the

[7] What is also called MAT definition, or Medial Axis Transform.

limitations of Blum's axes, devised more sophisticated algorithms, ones that promised better results (e.g., see the cow's axes computed by Feldman and Singh's Bayesian algorithm in Figure 3-4B (Feldman & Singh, 2006; Geiger et al., 2003; Katz & Pizer, 2003; Kimia, 2003; Ogniewicz & Kübler, 1995; Rezanejad et al., 2019; Rezanejad & Siddiqi, 2013; Siddiqi & Pizer, 2008).

How might the medial-axis perspective be compared with the proposal of object-centered representations built around a set of coordinate axes? One way to contrast these accounts is to consider the way in which the axes in each account accommodate shape representation. In the medial-axis view, the axes provide a *direct representation* of shape structures. Medial-axis researchers assume that the skeletal structure of a shape can be directly inferred from the structure of the medial axes. On this account, it is assumed that there exists a certain one-to-one correspondence between the axes' structure and the underlying structure (i.e., skeletal structure) of the objects. These assumptions are not necessarily true of axes in the coordinate-based view. Axes in the coordinate-based frame of reference *do not* provide a direct representation of shape structures. Rather, the axes provide bases for defining *spatial directions.* Spatial directions, in turn, are important for defining relationships among parts within an object. In the coordinate-axis viewpoint, the axes do not necessarily correspond to the structure of shapes in a one-to-one fashion. The frame of reference defined for the overall shape of the cow in Figure 3-4, for instance, may consist the principal axis parallel with the head-tail direction of the cow, and perhaps, another axis perpendicular with the principal axis (i.e., axis parallel with the dorsal-ventral direction). There is nothing in the structure of these perpendicular axes that corresponds to the cow's structure.

Given these considerations, it is not appropriate to compare these two accounts of shape axes on the basis of which account provides a better direct representation of shape structures. Recently, Ayzenberg and colleagues (2019), using a task pioneered by Firestone and Scholl (2014), asked a group of participants to 'tap anywhere they like' inside a simple line-drawn shape (e.g., a square, a trapezoid). The assumption of the study was that the participants had some covert knowledge about shape skeletons. By engaging them in a simple behavioral task where no explicit instruction concerning shape structures was given, the researchers assumed that the mental representations of the structures could somehow be revealed. The researchers found that participants' taps corresponded to locations along the medial axes but not the coordinate axes, thereby concluding that the medial-axis model provided a better account for shape representation. However, it was not clear what these results actually revealed about the coordinate axes. As coordinate axes do not directly represent shape structures, arguments for (or against) the coordinate-axis view could not be derived simply by asking what was the representation underlying shape structures. To argue for axes in the object-based frame of reference, it must be possible to demonstrate how the axes provide bases for defining directions within the object.

It is perhaps worth further discussing what it means to say that medial axes or a shape skeleton represents the shape of an object. Under the assumption that the brain analyzes object shapes in terms of parts, is each individual part of the object represented via a single skeleton or a set of skeletons? Or is the medial axis structure simply a way of capturing the shape's contour? Based on the current discussion in the literature it seems that the first view is more common. A frequently assumed benefit of medial axes is that the axis structure is constant over transformations of object parts, a property that is highly desirable in representing biological shapes (Blum & Nagel, 1978; Feldman & Singh, 2006; Firestone & Scholl, 2014). The hierarchical structure of medial axes for a body's structure, for example, does not change with movements of the body's appendages. This is due to the assumption that there exists a certain level of one-to-one correspondence between the axes and the parts. As mentioned, whether this one-to-one correspondence can be achieved in actual computer algorithms is a subject of ongoing research (Feldman & Singh, 2006; Geiger et al., 2003; Katz & Pizer, 2003; Kimia, 2003; Ogniewicz & Kübler, 1995; Rezanejad et al., 2019; Siddiqi & Pizer, 2008).

Regardless of the feasibility of one-to-one correspondence in medial axis computation, there clearly must be a way of defining spatial directions for relating object parts. It is possible that the visual system fulfills this requirement by computing perpendicular coordinate axes in addition to medial axes. That is, object shapes may be represented with respect to two types of axes each carrying fundamentally different types of information. Alternatively the visual system may construct coordinate axes based on medial axes and use them at different stage of shape processing (e.g., recognition vs. representation). Future research on shape axes may benefit by keeping both types of axes in consideration.

3.4. A Theory of Objects' Orientation Representation

The previous sections surveyed the basic function of axes in representing relationships of parts within an object. In this section, I discuss a theory that applies the notion of coordinate axes to explain how the brain represents the relationship between a whole object and the environment[8] (e.g., the orientation of a teapot in a room). The theory is particularly important for this thesis, as the details of the theory can be adapted to explain the mechanism by which the axes accommodate the representation of object structures. Additionally, the predictions of the theory provide bases for devising experiments for probing the axes' function.

3.4.1. Coordinate Orientation Representation (COR) Theory

Successful object recognition and interaction depend on the ability to correctly perceive and represent how the spatial structure of an object is oriented with respect to self or to the environment. For example, to reach out and safely grab a knife on a table, one must have access to visual information about how the knife is tilted and how its parts or features are arranged with respect to the body or to the hand (e.g., a knife is tilted approximately 90˚ relative to the body midline; the sharp tip of the knife points towards the right palm). A number of researchers assumed that the ability to compute the orientation of an object requires the visual system to somehow relate an object-centered frame of reference (e.g., frame defined by the knife) to an extrinsic frame of reference (e.g., the frame defined by the observer's body or the observer's hand) (Davidoff & Warrington, 2001; Harris et al., 2001; Ittelson et al., 1991; Priftis et al., 2003; Riddoch & Humphreys, 1988; Turnbull & Mccarthy, 1996).

According to McCloskey and colleagues (Gregory & McCloskey, 2010; McCloskey, 2009; McCloskey et al., 2006), orientations may be represented by relating a set of axes defined for an object's overall structure, to a set of axes defined by the observer or the environment (see Figure 3-5A for an example). In the Coordinate

[8] Or the observer

Orientation Representation (COR) theory, frames of reference are characterized in terms of perpendicular axes, and relationships between reference frames as relations between the sets of perpendicular axes. Figure 3-5A illustrates an example of the orientation representation of a knife. The COR theory was originally proposed to account for orientation in a 2D plane. In the example shown in this figure, the knife's frame is illustrated with two perpendicular axes: the object principal axis (i.e., Axis P) and the object secondary axis (Axis S).[9] The axes also have poles, which are defined on the basis of the structural features (or polar features) of the knife (e.g., the pointy tip of the knife, the curved edge of the blade). Likewise, the extrinsic frame is illustrated with two perpendicular axes: The vertical and the horizontal axes (i.e., Axis V and H). Here, generic features of the environment and of the observer's body (i.e., up, down, left, right) are selected to exemplify the poles of the extrinsic axes.

[9] The COR theory can be easily modified to account for 3D orientation representation, by defining three perpendicular axes for the object's frame and the extrinsic frame.

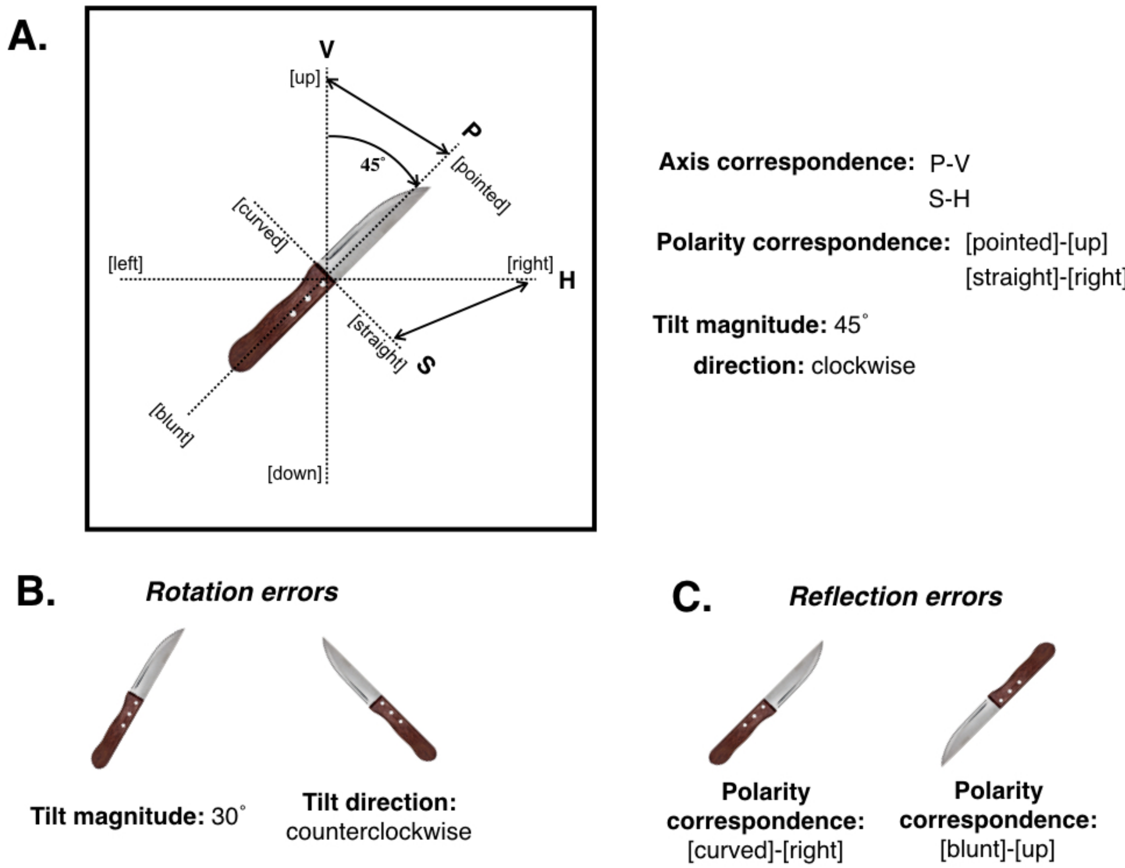

Figure 3-5. *(A) COR diagram for representing the orientation of a knife. (B&C) Examples of orientation errors predicted when certain COR parameters are misrepresented.*

In the COR theory, relationships between reference frames are represented via three parameters which specify different aspects of the relationships. The parameter *axis correspondence* specifies how axes from different frames are related to one another. In the example Figure 3-5A, the object principal axis is mapped to the vertical axis of the extrinsic frame, and the secondary axis to the horizontal axis (i.e., P-V, S-H). The mapping of the axes allows the other two parameters to be specified: The parameter *tilt* specifies how two axes are tilted relative to one another (e.g., in Figure 3-5A, the axes' tilt is 45˚ clockwise[10]), and the parameter *polarity correspondence* specifies how the axes' poles are mapped to one another. In Figure 3-5A, the *[pointed]* pole of the principal axis is mapped to the *[up]* pole of the vertical axis; and the *[straight]* pole of the secondary axis to the *[right]* pole of the horizontal axis.[11]

A critical assumption of the COR theory is that all the parameters play important roles for defining an object's orientation. In an experimental task where an observer attempts to remember the orientation of an object they previously saw (e.g., the orientation of a knife on the table), failures in encoding or retaining different COR parameters lead to different types of orientation errors. If the *tilt* parameter is misrepresented, then a "rotation error", or an error where the response is related to the target via a picture-plane rotation (Figure 3-5B), is expected.[12] If the *polarity correspondence* is misrepresented, then a "reflection error", or an error where the response mirror-reflects the target across some intrinsic axis of the object (Figure 3-5C), is predicted. Figure 3-5C illustrates two examples of reflection error that take place across the principal axis (i.e., left side of the figure), and the secondary axis (i.e., right side) of the knife. The reflection error across the object principal axis occurs due to an incorrect mapping of axes' poles between the object secondary axis and the extrinsic

[10] In the diagram, the tilt information is displayed only between one pair of axes (i.e., Axis P and V). One can imagine that the same tilt information is specified for the other pair of axes (i.e., Axis S and H). The COR theory makes no prediction regarding whether the tilt parameter is redundantly coded for both pairs of axes, or only for one pair.

[11] Here too, the polarity correspondence may in theory be redundantly encoded for both ends of an axis. For example, for the P-V pair, the mapping may be specified not only for the [pointed]-[up] features of the axes, but also for the [blunt]-[down] features.

[12] In Figure 3-5B if the magnitude of tilt is represented as being 30˚ instead of 45˚, the knife's orientation may be remembered as having a slightly incorrect tilt (Figure 3-5B, left). If the direction of the tilt is encoded as 'counterclockwise' instead of 'clockwise', the knife may be remembered with a correct amount of tilt, but with the tilt inclining in the different direction with respect to the vertical axis of the extrinsic frame (Figure 3-5B, right)

vertical axis: Instead of the *[straight]* pole, the *[curved]* pole of the knife's axis is mapped to the *[right]* pole of the extrinsic axis. The reflection error across the object secondary axis occurs because of an incorrect polarity mapping between the object principal axis and the extrinsic horizontal axis: The *[blunt]* pole instead of the *[pointed]* pole, is mapped to the *[up]* pole of the extrinsic vertical axis.

3.4.2. Function of Axes in The COR Theory

In the COR theory, the function of axes is similar to what I have been discussing for shape representation. Axes in COR provide a basis for representing relationships among parts, and for representing the location and orientation of the object by relating it to an extrinsic frame of reference.

Figure 3-6 illustrates this point by removing one object axis (i.e., the secondary axis of the knife) from the COR representation. As only one object axis is available (i.e., the principal axis), there is no basis for encoding relations of object features along a direction not parallel with the available axis. Specifically, the curved vs. the straight side of the knife cannot be distinguished in terms of where they are along the (non-existent) secondary axis. As a result, it is not possible to define whether the curved (or the straight) side of the knife is facing leftward vs. rightward, or upward vs. downward. In COR vocabularies, there exists no basis for relating the polar features of the object secondary axis to the features of the extrinsic axes.[13] Thus, the two orientations of the knife shown in the figure cannot be differentiated.[14]

---

[13] In addition, if the blade and the handle of the knife are treated as separate parts, there is no way to represent that the straight part of the handle is on the same side as the straight part of the blade.
[14] These orientations correspond to mirror-reflections across the object principal axis.

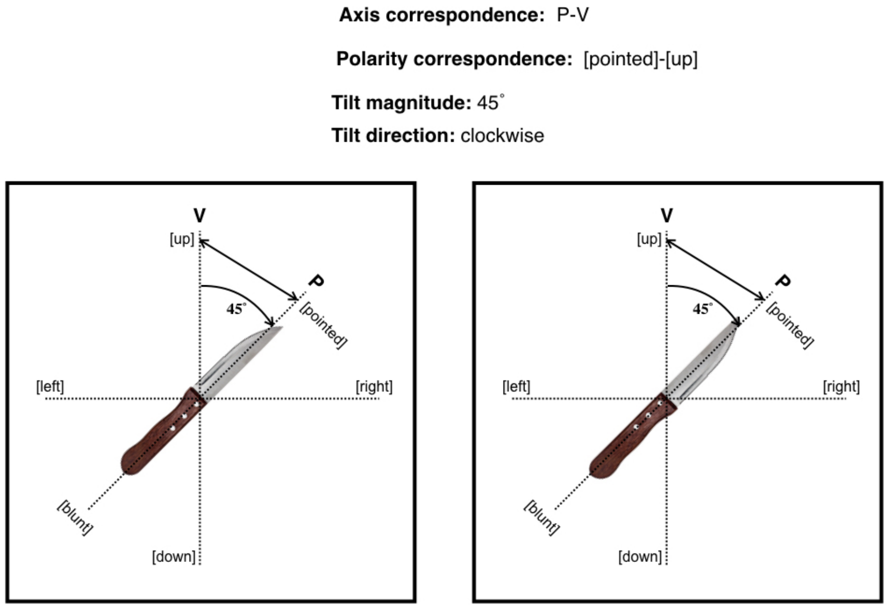

Figure 3-6. *Representations of orientation where only the principal axis is present for the object's reference frame. In the COR framework, certain orientation differences cannot be distinguished.*

Another way in which axes play important roles in the COR theory is that they provide bases for defining spatial directions for encoding the tilt information. To encode the tilt of an object, it must be possible to determine which direction the object is tilted. The axes of the extrinsic reference frame supply a basis for defining the directions. In Figure 3-6 above, for example, the knife is tilted in the clockwise direction--that is, with its pointed feature pointing from the up pole of the vertical axis towards the right pole of the horizontal axis. If any one axis of the extrinsic frame is missing (e.g., the horizontal axis, see Figure 3-7 below), then it is no longer possible to specify the complete tilt information. In the figure below, it is not possible to differentiate the directions in which the object is tilted.

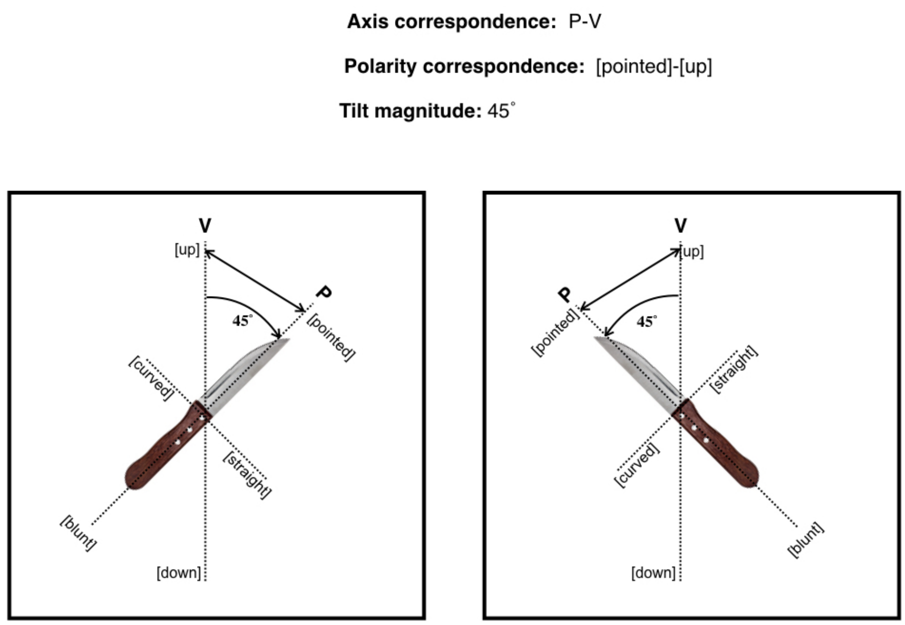

Figure 3-7. *Representations of orientation where only the vertical axis is defined for the extrinsic reference frame. In the COR framework, it is not possible to distinguish certain tilt differences.*

3.4.3. Orientation Recall Experiment

Empirical evidence that provided support for the COR theory came from a series of studies where participants were asked to remember the orientation of an object they saw several seconds before (Chaisilprungraung et al., 2019; Gregory et al., 2011; Gregory & McCloskey, 2010). In an orientation memory task, participants saw a photograph of a target object at a particular orientation, such as the comb shown in Figure 3-8, and attempted shortly thereafter to recall or recognize the object's orientation. The researchers found that when participants made an error in this task, their orientation error often took the form of a reflection across the object principal axis. In these experiments, this type of error was referred to as the Object Principal Axis or OPA reflection error. Figure 3-8 shows a plot of the error distribution reported for one of the experiments in the study by Gregory and McCloskey (Gregory & McCloskey, 2010). The OPA error clearly occurs at a rate disproportionately higher compared to other types of errors.

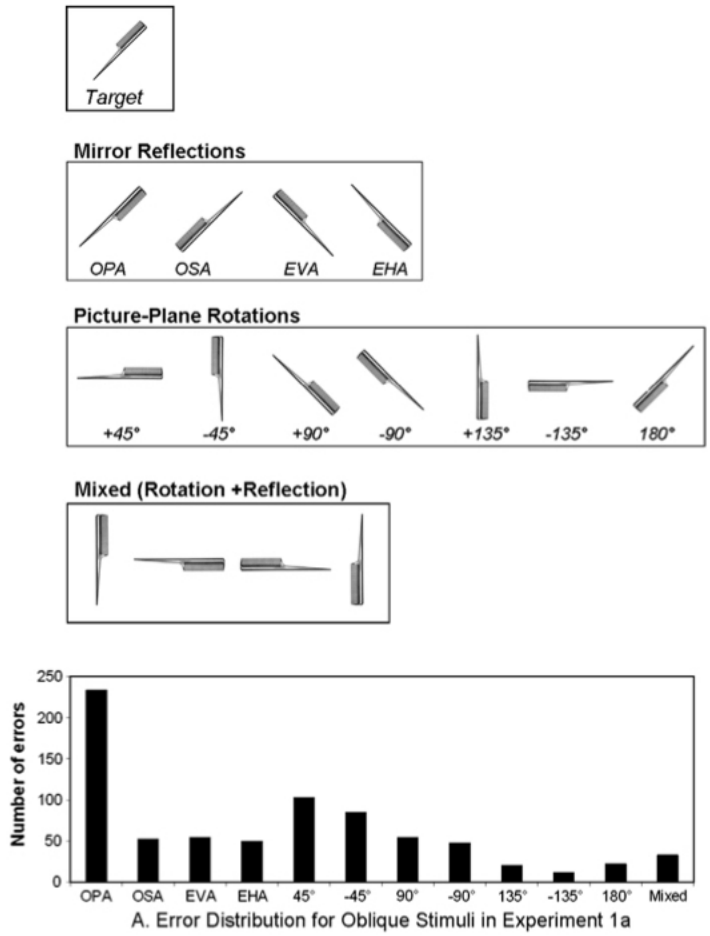

Figure 3-8. *Plot of error distribution of an orientation recall experiment (Gregory & McCloskey, 2010), along with illustrations of what each error type corresponds to.*

The high frequency of the OPA errors could be explained in the framework of the COR theory. Specifically, this type of error might frequently occur because certain features along the object secondary axis (e.g., the curved side vs. the straight side of the knife) were perceptually similar. Due to the perceptual similarity, remembering how the polar features of the secondary axis were related to features of the environment might be particularly difficult. Errors involving reflections across the object principal axis likely occurred because of a misrepresentation of the axes' polarity correspondence (Gregory et al., 2011; Gregory & McCloskey, 2010) (see also Figure 3-5C).

3.5. Evidence for Perpendicular Coordinate Axes in Shape Representation

In the proposal thus far, I argued that relationships among parts of an object must be defined with respect to multiple axes. If object shapes are defined in 2D, then two axes are required, and if shapes are defined in 3D, three axes are required. Is it possible to obtain evidence that multiple axes are indeed present in the object's reference frame?

A type of evidence supporting the notion of multiple axes can be hypothesized on the basis of the COR theory. The COR theory posits that orientation errors like the OPA reflections occur because some failure(s) occur in encoding or retaining the *polarity correspondence* parameter (see Figure 3-5C). Specifically, due to an incorrect mapping of axes' polar features between the object's reference frame and the extrinsic frame, an error occurs that involves an object's reflection across an axis of its intrinsic frame of reference. If it is somehow possible to demonstrate that errors of this kind frequently take place across more than one axis--that is, not only across the object principal axis, but also across some other axis (or axes) not parallel with the principal axis--then it is possible to obtain evidence for multiple axes. In other words, the orientation recall experiment can be adapted for the purpose of inferring object axes from the axes of reflection.

This was precisely the aim of my recent study (Chaisilprungraung et al., 2019). In each trial of the experiment participants saw an object stimulus presented at an orientation on the screen. The object's orientation was selected at random from 360 possible tilts ranging in 1˚ increments from 0˚ to 359˚, and also two enantiomorphs (i.e., mirror-reversals). Participants were instructed to remember the object's orientation. Shortly after the target presentation, they saw a probe stimulus (i.e., the same object appearing at a different random orientation), and adjusted the orientation of the probe (i.e., by rotating or flipping the object) to match it to the orientation of the previously-seen target.

When an orientation reflection error was made in the experiment, I performed an analysis to identify the axis of reflection (see Figure 3-9A), and then plotted the reflection axis relative to a pre-chosen vertical orientation of the target (Figure 3-9B). Across multiple trials of the experiment, the axes of reflection were overlaid in the same plot and analyzed (Figure 3-9C).

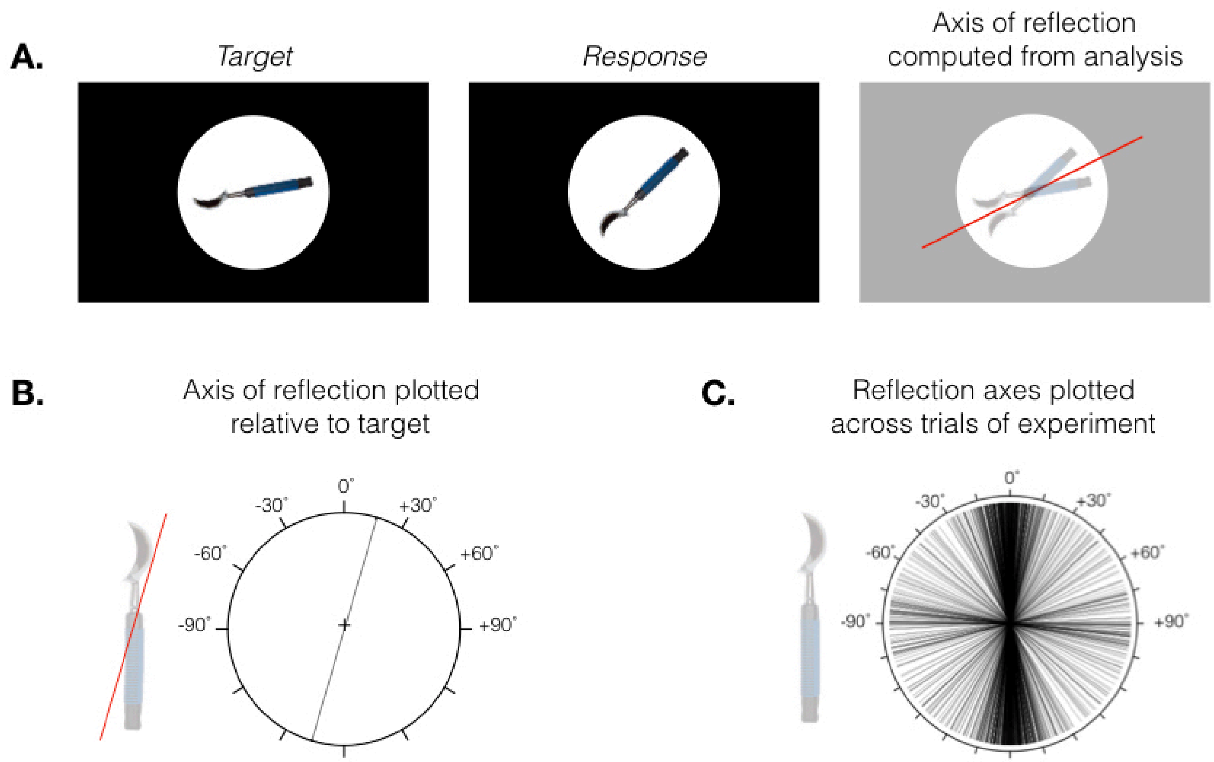

Figure 3-9. *(A) Illustrations of how an axis of reflection is computed and (B) displayed. (C) A plot showing all axes of reflection of the study (Chaisilprungraung et al., 2019)*

As shown in the figure above, the axes of reflection were mostly aligned with (or nearly aligned with) 0˚, which corresponded to the object's principal axis of elongation. The preponderance of reflections at 0˚ replicated the previous finding where the OPA errors were observed at high frequency. In addition to these reflection errors, however, we also observed another type of error that took place across an axis aligned with (or nearly aligned with) 90˚. The occurrence of these latter errors was taken as evidence for a secondary object axis perpendicular to the principal axis.[15] In short, the results

[15] Formal analyses verified that reflections across the secondary axis (i.e., OSA errors) occurred more frequently than expected by chance. The analyses indicated that axes of reflection in the range 90˚ $\pm$ 15˚ (a range assumed to correspond to the range of the object secondary axis) were significantly more common than axes in the 'baseline' range (i.e., the range pre-defined as 30˚ to 60˚ and -30˚ to -60˚) (see Chaisilprungraung et al, 2019 for more details).

suggested that more than one axis is present in the object's frame of reference. The success of the experiment also hinted at the usefulness of the COR theory as a basis for investigating the object axes.

3.6. A Potential Framework for Probing Axes' Function

The COR theory also provides a potentially useful framework for theorizing about and testing the axes' function. Through discussions in this chapter, I argued that object axes provided bases for defining spatial directions. The intrinsic directions in the object-centered frame of reference are useful in some way for relating parts within an object (e.g., where the teapot's spout is and how it is tilted relative to the teapot). What might be the mechanism for relating object parts?

The COR theory provides a potential answer. To represent spatial relations among parts of an object, the visual system might encode different parameters for relating a set of axes defined for different object parts. In this thesis, I assume that object parts are represented in a *hierarchical* manner, where parts at a more local level are always defined relative to parts or structures at a more global level (Hoffman & Richards, 1983; Marr, 1982; Palmer, 1977). The sections below discuss my proposals about how the COR theory can be adapted to explain the encoding of object parts' orientation and location.

3.6.1. Representation of Relative Part Orientation

The COR mechanism which involves specifying three parameters (i.e., axis correspondence, polarity correspondence, and tilt) may be straightforwardly adapted to account for the representation of relative part orientations. Figure 3-10 illustrates this idea using a simple artificial object. Here, the orientation of the small object part is represented with respect to the large object part. In this diagram upper-case letters (i.e., P, S) are used for denoting the principal and the secondary axes of the large object part, and lower-case letters (i.e., *p, s*) for axes of the smaller part. In addition, positive and negative signs (i.e., $+, -$) are used to denote the axes' polar features. For simplicity the polar features of the large part's axes are denoted with white-centered signs, and the features of the small part's axes black-centered signs.

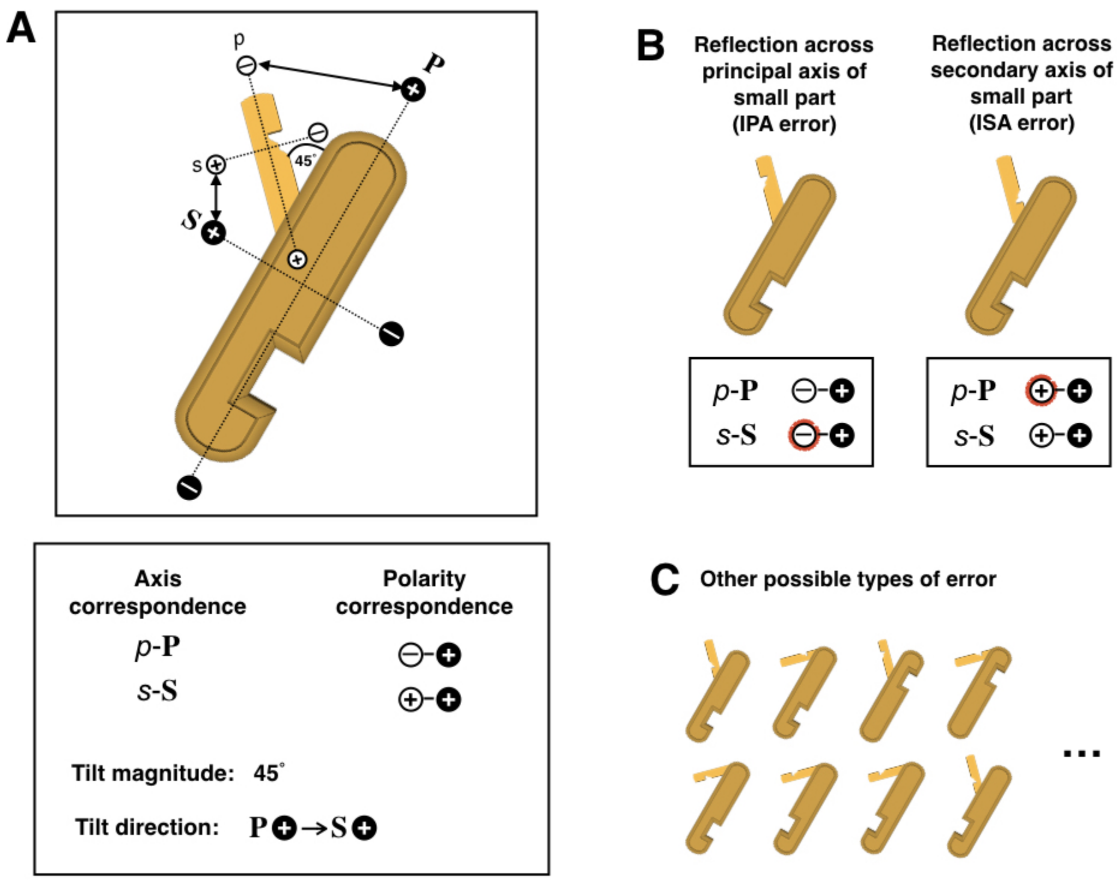

Figure 3-10. *(A) Illustration of the COR theory adapted for accounting for the representation of object part relations. (B) Errors involving reflections of object parts which occur as a result of a misrepresentation of polarity correspondence. (C) Other possible types of errors which may be expected in a shape recall task.*

Relations between the systems of perpendicular axes are defined by establishing three COR parameters. For the present example, the *axis correspondence* is established by mapping the parts' principal axes together (i.e., *p*-P), and the secondary axes together (i.e., *s*-S). The *polarity correspondences* are established by mapping the [–] feature with the [+] feature for the principal-axis pair, and the [+] feature with the [+] feature for the secondary-axis pair. The *tilt* of the axes is defined based on the amount and the direction of deviation of Axis *p* from Axis P. In the example of the figure Axis *p* is tilted from Axis P for 45˚. This tilt amount must be defined based on intrinsic directions given by the axes (i.e., Axis P and Axis S) of the global reference frame. Here I define intrinsic directions by referring to the polar features of Axis P and Axis S: The tilt direction of Axis *p* is consistent with the direction of the shortest distance taken from the [+] feature of Axis P to the to the [+] feature of Axis S.[16]

If some failure(s) occur in the encoding or retaining of the parameters, then certain types of orientation errors are expected. Figure 3-10B illustrates examples of errors that would occur if the polarity correspondence parameter is forgotten. If participants incorrectly remember how the polar features of are related across the secondary axes of the object part (see Figure 3-10B, left), then they will make an error where the small object part is reflected across its intrinsic principal axis (henceforth: IPA or Intrinsic Principal Axis reflection error). If they misremember how the features are related across the object principal axes (Figure 3-10B, right), then the error will correspond to a reflection across the object part's intrinsic secondary axis (henceforth:

[16] Directions such as 'clockwise' and 'counterclockwise' may not be used here because these directions are defined based on some features of the extrinsic frames (e.g., [up], [down], [left], [right]).

ISA or Intrinsic Secondary Axis reflection error). Errors involving a rotation of the small object part relative to the overall object can also occur if some aspects of the tilt information (i.e., tilt magnitude and tilt direction) are misremembered.

The COR theory and its error predictions provide useful bases for probing the function of axes in shape representation: If the mechanism for relating parts within an object is indeed the same as the mechanism for relating the overall object to the environment (or the observer), then similar patterns of errors are expected. In previous studies where participants' memories of objects' orientations were tested, a significant portion of the errors corresponded to reflections across an object intrinsic axis (i.e., OPA and OSA error) (Chaisilprungraung et al., 2019; Gregory et al., 2011; Gregory & McCloskey, 2010). In a novel task where participants are asked to remember the structure of stimulus objects similar to that shown in Figure 3-10, we may expect significant rates of errors in the form of reflections of the small object parts across their intrinsic axes (i.e., IPA and ISA errors).

3.6.2. Representation of Relative Part Location

Besides the representation of parts' orientations, the COR theory can also be adapted to explain how the relative locations of object parts are represented. A possible way of representing the location of a local object part relative to a global part is by encoding the distance between the *origins* of the parts' axes (i.e., the point intersected by the axes; see red circles in Figure 3-11A). In the example below, the origin of a part's axes corresponds to the point where the principal axis and the secondary axis meet. For convenience, the polarities of these axes[17] are left out from the illustration, as the current

[17] i.e., what were formerly shown as black-centered [+] and [-].

proposal assumes that only the axes' origin is necessary for representing the object part's location.

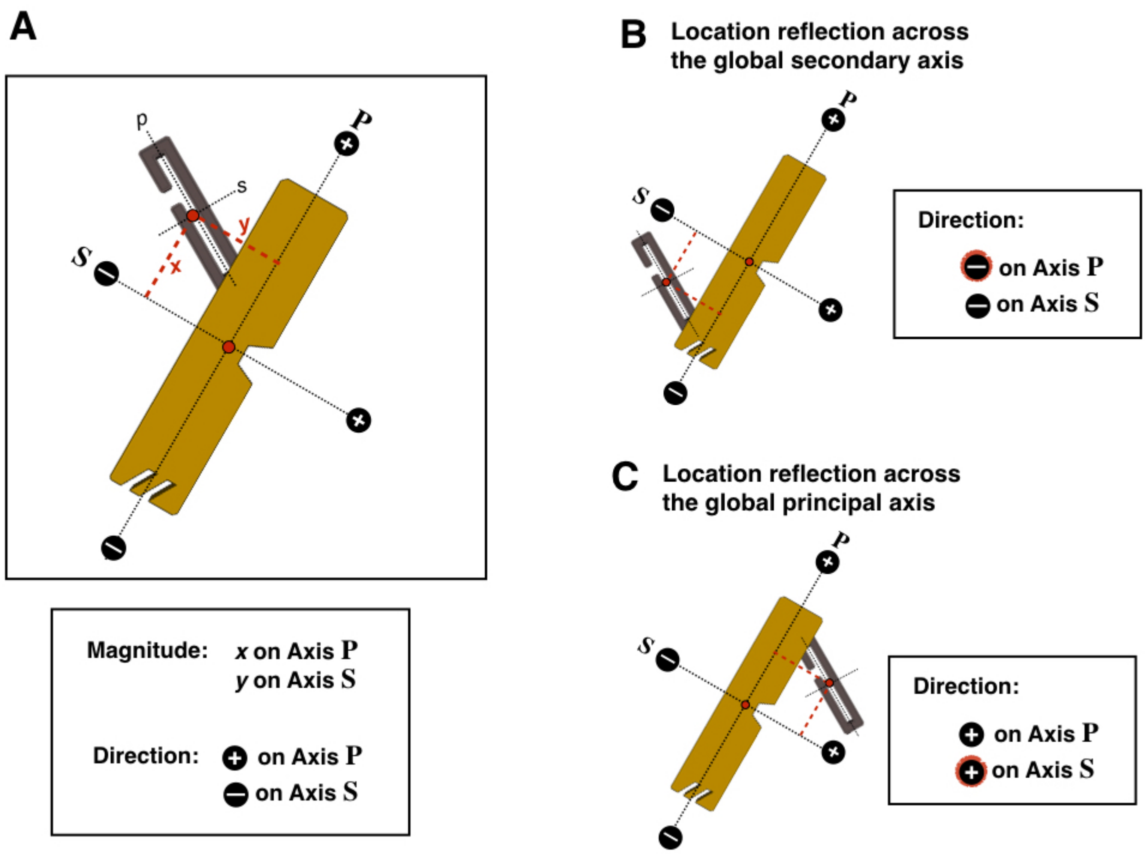

Figure 3-11. *(A) A possible coordinate-based model for representing the relative location of object parts. (B&C) Errors involving location reflection which may be expected to occur if certain parameter (i.e., direction) of the representation is incorrect.*

Like orientation representation, the representation of object parts' locations may be carried out by specifying different parameters. The amount of distance between the axis origins (see *'x'* and *'y'* in the figure) can be represented by specifying a parameter for the distance's *magnitude.* The directions of the distance (i.e., whether the small part is nearer to one side of the base part or the other) can be represented by specifying a parameter about *direction.* For the present example, the small object part's distance from the middle of the base is ‘x’ towards the [+] pole of Axis P, and ‘y’ toward the [-] pole of Axis S.

Different types of location errors may be expected if some parameter(s) of the representation are forgotten. If the magnitude of the distance is misremembered (e.g., due to some noise in the memory), then the error will involve slight imprecision in the location of the small object part relative to the base. If the direction is misremembered, the error will correspond to a reflection of the small part’s location across an axis of the base. Figure 3-11B&C illustrates two examples where the small part is remembered as being on the wrong side or the wrong end of the base. If the direction on the Axis P (i.e., Figure 3-11B) is incorrectly encoded, then the small part may be remembered as being closer to the wrong end of the base (i.e., being nearer to the end resembling the comb’s teeth, rather than the flat end). If the direction of distance along Axis S (i.e., Figure 3-11C) is misrepresented, the small part may be remembered as being on the side of the base with a notch in the middle.

It is important to note that the expected pattern of location reflection errors depend on an important assumption about how the origin of the coordinate axes are defined. In Figure 3-11, the origin of the axes of the small object part corresponds to the part’s center. However, another way of defining the axes’ origin is possible: the origin may be defined so that it corresponds to the place where the parts are attached (e.g., see Figure 3-12). If this is the case, a different pattern of location reflection error is expected. Specifically, the error should preserve the distance from middle of the place of attachment (see distance *‘x’* in Figure 3-12), instead of the distance from middle of the part’s center.

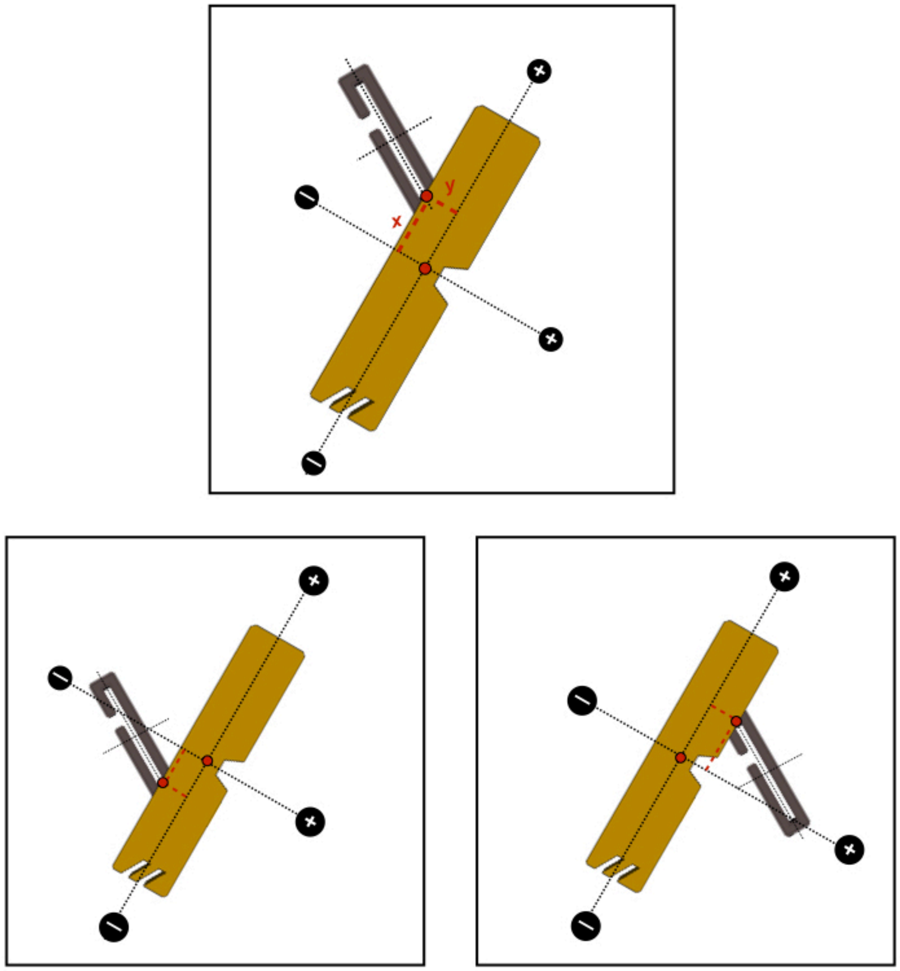

Figure 3-12. *Alternative way of defining the origin of the small part's axes (top), and the resulting location reflection errors (bottom).*

In the first two experiments of this thesis, I discuss procedures for probing object shape representation, which involves asking participants to recall the structure or the orientation of an object they previously saw. Evidence obtained from these experiments will be important not only for determining the relevance of the COR theory in shape representation, but also for probing how origins of object parts' axes are defined. Questions about how origins of *whole-object* axes are defined will be the focus of the latter portion of this thesis (Chapter 4; Experiment 3 & 4).

CHAPTER 4. GEOMETRIC CONSTRAINTS

Chapter 3 of this thesis dealt with questions related to the *function* of axes in coordinate-based shape representation. In chapter 4, I considered questions about how the visual system determines object axes given a shape. That is, what geometric properties of shapes constrain how axes are defined for shapes? As mentioned previously, most psychological studies assumed that there existed a single most important axis (i.e., the principal axis) defined for an object-centered reference frame (see § 3.2. Axes in Shape Representation). In this literature, questions concerning *geometric constraint* of axes were posed in these terms: What abstract geometric properties determine how the principal axis is defined for a shape?

Based on collective findings, researchers generally agreed that the principal axes were defined in parallel with either shape elongation or shape symmetry (Large et al., 2003; Leek & Johnston, 2006; Ling & Sanocki, 1995; McMullen & Farah, 1991; Morikawa, 1999; Quinlan & Humphreys, 1993; Rock, 1973; Sekuler & Swimmer, 2000; Smith et al., 2014; Tarr & Pinker, 1990). This conclusion was based on results from various experimental paradigms. In the first section of this chapter (§ 4.1. Existing Evidence for Shape Elongation), I provide a short review of these studies. In the second section (§ 4.2. A Closer Look at The Notion of Elongation), I further examine the notion of 'shape elongation', pointing out a rarely-recognized fact that the shape elongation can be defined in more than one way in some stimulus objects. An experiment conducted in my previous work bearing evidence dissociating the two construal of shape elongation is discussed in this section. In the final section (§ 4.3. Origins of Coordinate Axes), I discuss questions about how origins of coordinate axes are defined. The discussion in this section provides a basis for two experiments discussed in Chapter 5.

4.1. Existing Evidence for Shape Elongation

Among the earliest research on geometric constraints was a study by Quinlan and Humphreys (1993). In their study participants were asked to draw a line they thought went most naturally with a simple geometric shape, and a shape's principal axis was inferred from participants' drawings (see also discussions in § 3.2. Axes in Shape Representation). Quinlan and Humphreys were interested in comparing shape elongation and shape symmetry, thus their stimuli were selected so that each possessed the axis of elongation only (i.e., *Elongated* condition; Figure 4-1A), the axis of symmetry only (i.e., *Symmetric*), both axes in parallel (i.e., *Conjoint*), both axes in perpendicular (i.e., *Disjoint*), or neither axis (i.e., *Asymmetric*). Shape elongation was defined on the basis of the longest line that could be drawn across a shape, whereas shape symmetry, the line that bisects a shape into two identical halves (Figure 4-1A).

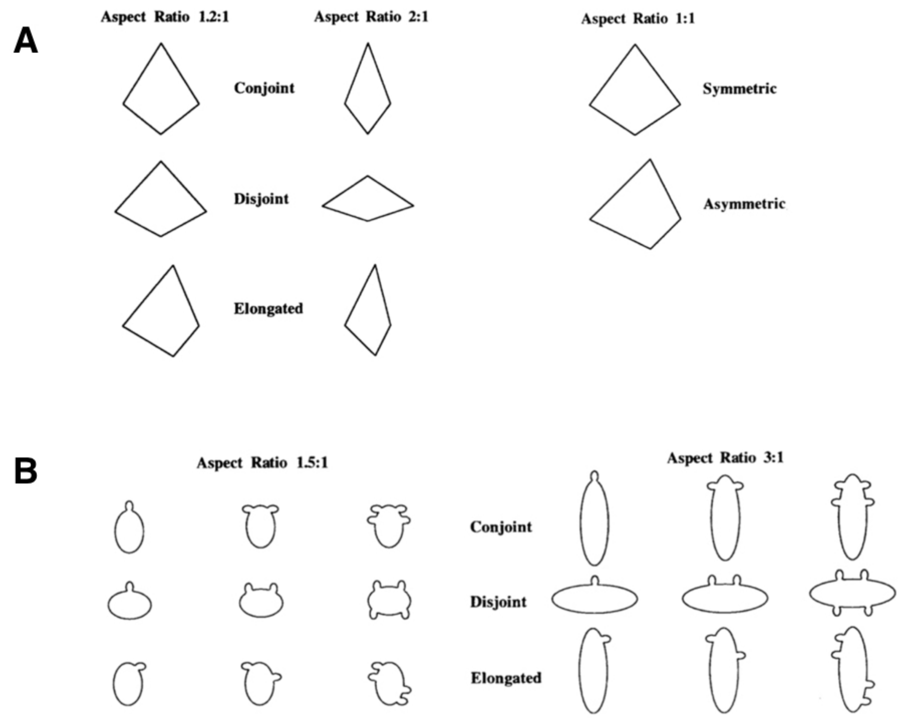

Figure 4-1. *(A) Stimuli from Quinlan and Humphreys (1993's experiment), and (B) stimuli in Sekuler and Swimmer (2000)*

Quinlan and Humphreys found that whenever an axis of symmetry was present in a shape (i.e., *Conjoint*, *Disjoint*, and *Symmetric* stimuli), participants tended to draw a line parallel with that axis. Whenever the axis of symmetry was absent (i.e., *Elongated* and *Asymmetric*), participants' responses were evenly divided across the two diagonals of a shape (i.e., the two perpendicular lines that could be drawn across the opposing vertices of a shape).[18] Based on these results, they suggested that in a coordinate-based shape representation, only shape symmetry, and not shape elongation, was important in constraining how the principal axes were defined.

[18] All stimuli in the experiment were displayed only at the orientations where the opposing vertices were aligned with the vertical or the horizontal axis (e.g., Figure 4-1). Thus, participants' drawings always corresponded to either a vertical or a horizontal line.

In a later study, Sekuler and Swimmer (2000) suggested an alternative interpretation for Quinlan and Humphreys' finding: The reason why the participants did not tend to select the axis of elongation might be because the elongated stimuli (i.e., *Conjoint, Disjoint, Elongated*) were never tested at an aspect ratio high enough.[19] In an adaptation of the original study, Sekuler and Swimmer replaced the polygon-shaped stimuli with stimuli that were made up of a large ellipse and auxiliary bumps placed at various locations along surfaces of the ellipse (Figure 4-1B). They increased the aspect ratio of these circular-shaped elongated stimuli[20] , and found that as the aspect ratio increased, participants were also more likely to select the axis of elongation. The study therefore suggested that shape elongation, in addition to shape symmetry, played important roles in defining coordinate axes.

Other evidence for the saliency of elongation and symmetry was obtained through somewhat subtler task paradigms (Large et al., 2003; Ling & Sanocki, 1995; McMullen & Farah, 1991; Morikawa, 1999; Sekuler, 1996; Smith et al., 2014; Tarr & Pinker, 1990). Perhaps noteworthy among these, Morikawa (1999) demonstrated that the perception of direction of a moving shape was biased by the presence of the axis of elongation, as well as the presence of the axis of symmetry. Participants viewed a shape moving on screen along a certain linear path. For an elongated or a symmetric shape (i.e., the latter three columns in Figure 4-2A), this trajectory of motion could be parallel with or oblique to the shape's axis of elongation or symmetry (Figure 4-2B). Participants indicated what they perceived to be the moving trajectory.[21] Morikawa found that when a stimulus was

[19] The maximum aspect ratio of the elongated stimuli was 2:1, see Figure 4-1A.
[20] The maximum aspect ratio in Sekuler and Swimmers' study was 6:1.
[21] In particular, they saw a probe arrow displayed at a corner of the screen, and adjusted the arrow so that it pointed in the same direction as the perceived trajectory. In another experiment, they

elongated or symmetrical, the direction of motion was often perceived as being more aligned with the axis of elongation or the axis of symmetry than it actually was.

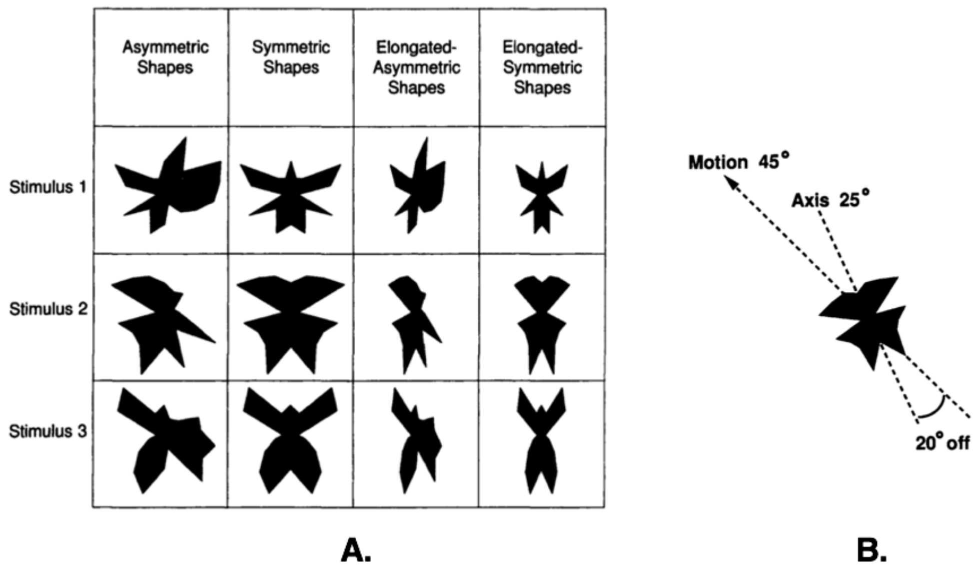

Figure 4-2. *(A) Examples of stimuli from Morikawa (1999) (B) An example of a symmetric shape moving along a trajectory which diverged 20° from the axis of symmetry. The dotted line and arrow were not visible during the actual experiments.*

4.2. A Closer Look at The Notion of Elongation

The previous discussion suggests that shape elongation and shape symmetry are important geometric determinants for how object axes are defined. The notion of 'elongation', however, is not always easily defined in some stimulus objects. For an illustration of this point, consider the hatchet shown in Figure 4-3. The structure of the hatchet is made up of an elongated part (i.e., the handle) and a protruding part (i.e., the blade). How might the principal axis of elongation of this type of object be defined?

---

saw a shape moving along a trajectory, and then disappearing a few seconds right before it reached the final destination. Participants placed a dot on the screen at the location where they expected to be the intended final destination.

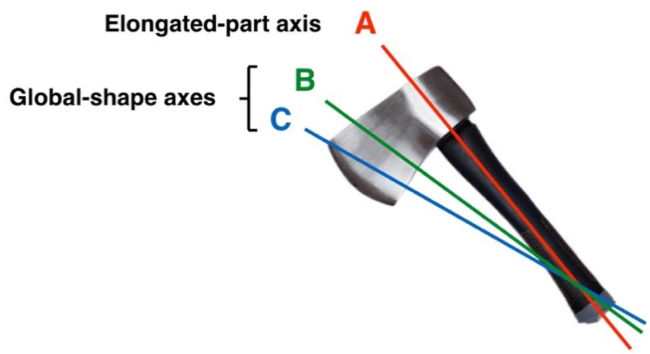

Figure 4-3. *Illustration of a hatchet showing three potential definitions of an object's axis of elongation: (A) the long axis of the most elongated part, (B) the axis of least second moment, and (C) the longest span axis.*

One possibility is that the principal axis is defined in parallel with the object's longest part (i.e., the handle, Axis A in Figure 4-3). For simplicity I refer to this hypothesis as the Elongated-Part Hypothesis. Another possibility is that the principal axis is defined in some way that takes into account the overall elongated shape of the object (i.e., the Global-Shape Hypothesis). As the hatchet is made up of the handle and the blade, a global-shape axis must tilt so that it extends across both of these object parts in some ways. A relatively crude way to operationally define this axis is to find two points on an object shape that are the furthest. That is, a global-shape axis may correspond to the longest line that can be drawn over a shape (i.e., the Longest-Span Axis, see Axis C in the figure; see also discussions in the previous section about how Quinlan and Humphreys defined shape elongation). Another, perhaps more reliable definition a global-shape axis is to find the axis that 'best fits' all points on a shape. The axis of least second moment (i.e., Axis B in the figure) corresponds to the axis that yields the minimum sum of least squared distance to all points in a shape.[22] In the computer vision research, this axis is the preferred definition of shape elongation[23] (Haralick & Shapiro, 1991).

It is worth stressing that the important question here is not how terms such as 'elongation' or 'global-shape axis' should be defined. The important question is which aspects of shape geometry are relevant for determining object axes. As previously mentioned, different ways of defining the axes have important theoretical implications for how object parts' relations are represented. Predictions for types of errors in a shape memory task may also critically depend on assumptions about how object axes are defined (see § 3.6.2. Representation of Relative Part Location). Understanding of geometric constraints of shape axes are important for theorizing about shape representation.

In my previous study where I adapted the orientation recall procedure to explore the nature of object axes (Chaisilprungraung et al., 2019; see § 3.5. Evidence for Perpendicular Coordinate Axes in Shape Representation), I also evaluated the elongated-part and the global-shape hypotheses. Participants were asked to recall the orientation of objects that were made of an elongated part and a protruding part, like the hatchet in Figure 4-4 (for a complete set of stimuli, see Figure 5-31). The method for inferring object axes was the same as previously mentioned. Participants viewed a target object at the center of the screen (Figure 4-4A). After the target disappeared, they attempted to adjust the probe object (i.e., by rotating or flipping it) to match the target. When a

[22] This computation is equivalent to that used for finding the principal component in data analysis. In PCA (Principal Component Analysis), the principal component corresponds to the direction that explains the highest variance in the data, given its distribution.
[23] And also the preferred definition for 'principal axis'.

reflection error was made, I computed the axis of reflection and plotted it relative to the target stimulus (Figure 4-4B). The elongated-part hypothesis predicted that a significant portion of reflection errors would occur across an axis parallel with the object's longest part, and also across the axis perpendicular to the elongated part (i.e., the axes that, if displayed in the circular plot, would have tilts corresponding to 0˚ and 90˚). The global-shape hypothesis predicted that the tilts of the reflection axes would be determined somehow by the overall elongated shape of the object (i.e., axes with tilts similar to the tilts of the longest-span axis, or the axis of least second moment).

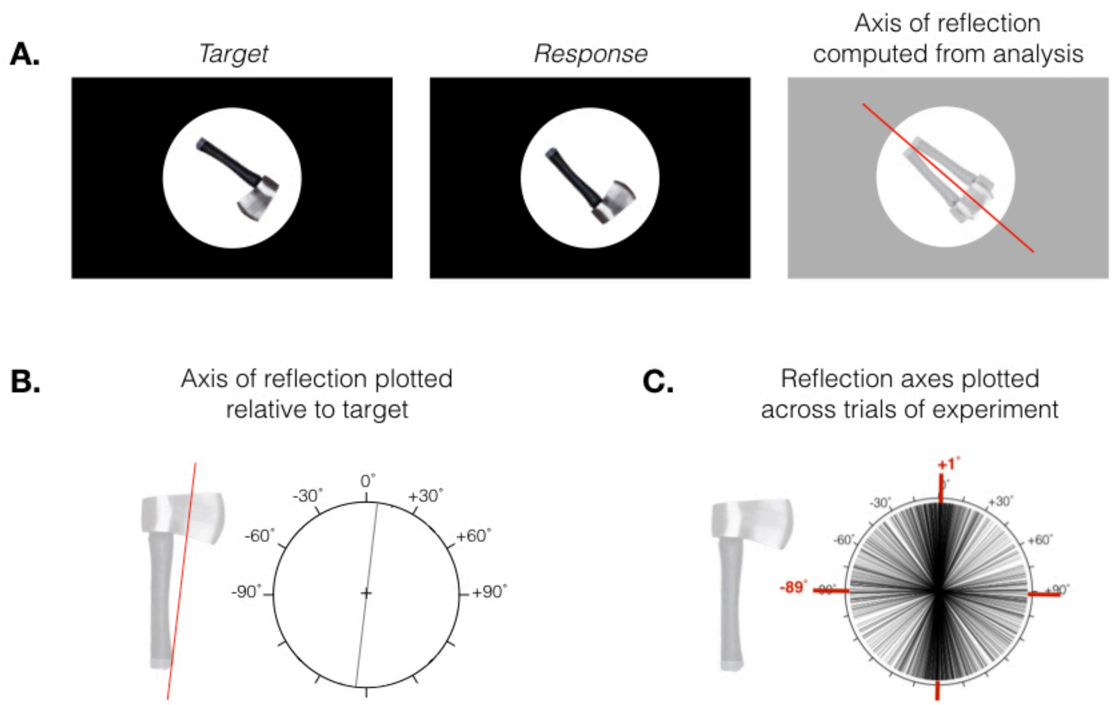

Figure 4-4. *(A) Illustrations of how an axis of reflection is computed and (B) displayed. (C) A plot showing all axes of reflection in the second experiment of the study by Chaisilprungraung and colleagues (2019).*

The result of the experiment revealed that most axes of reflection were parallel or near-parallel to the elongated part (i.e., in Figure 4-4C, these axes are displayed in a cluster around 0˚). To some extent we also observed axes of reflection perpendicular to the elongated part (i.e., axes clustering around 90˚ in the figure).[24] To measure the overall orientations of the axes, we performed an additional analysis to identify the pair of perpendicular axes that yielded the minimum sum of squared distance to all data points (i.e., the 'best-fit' axes, designated by the red ticks in Figure 4-4C). We found that the best-fit axes had the orientations of +1˚ and -89˚, which were almost perfectly aligned with the axes predicted by the elongated-part hypothesis.[25] The experiment's results therefore strongly suggested that object parts' relations are encoded relative to the axes defined by the elongated part.

4.3. Origins of Coordinate Axes

Although previous findings revealed interesting insights about the geometric determinants of shape axes, there is another important question about how axes are defined that cannot be easily answered based on data of these experiments. In my previous study I demonstrated that axes' tilts were determined on the basis of the object's elongated part. What might be the basis for determining the point where the axes intersect (i.e., the origin of the coordinate system)? Figure 4-5 illustrates two possible scenarios. In the top-row diagram, the origin of the coordinate axes corresponds to the center of the elongated part (henceforth: Elongated-Center Hypothesis). The object secondary axis defined in this way bisects the length of the elongated part, and the principal axis bisects the width of the elongated part. Reflections across axes of the

---

[24] Similar to the experiment with simple-elongated objects (e.g., scoop, see §3.5), these axes occurred more frequently than expected by chance, consistent with the notion of the perpendicular-axis reference frame.
[25] And nowhere near the axes predicted by the global-shape hypothesis.

Elongated-Center Hypothesis preserve the location of the elongated part on the screen.

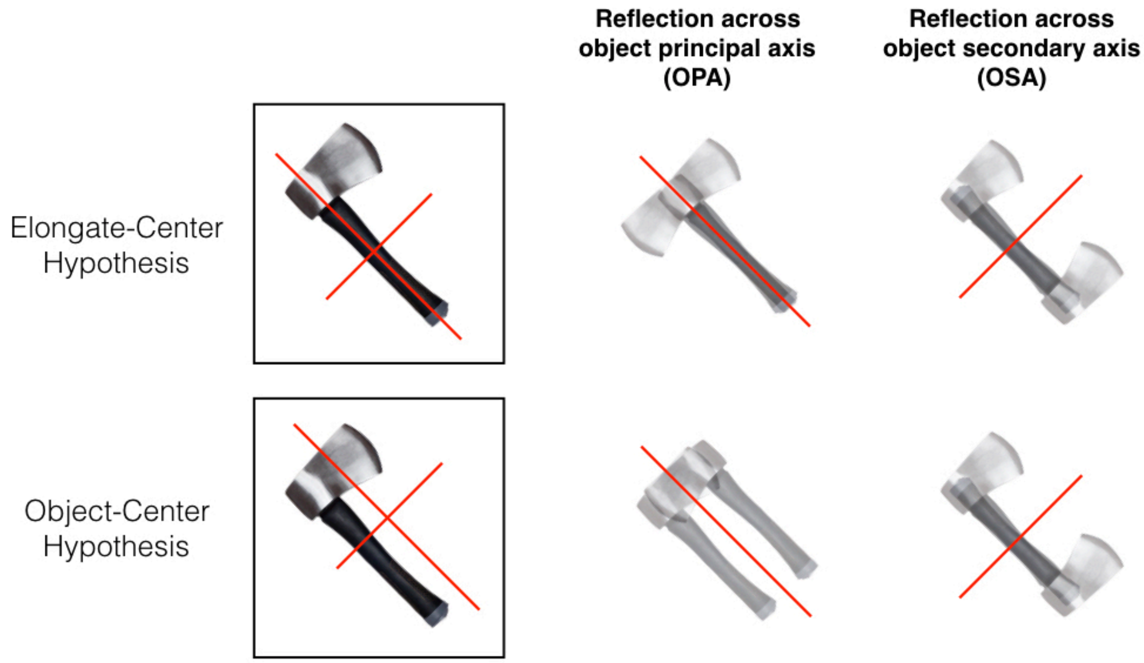

Figure 4-5. *Hypotheses concerning how origins of coordinate axes may be defined, and predictions for each hypothesis.*

Another possible way of defining the axes' origin is based on the center of the overall object (henceforth: Object-Center Hypothesis). Given a shape, there are many ways of deciding what the center of the shape should correspond to. In the bottom row of Figure 4-5, I illustrate an example where the middle corresponds to the center of the smallest bounding box that can contain the object.[26] The object secondary axis defined in this way bisects the object's length, and the principal axis bisects the object's width. Reflections across the principal axis (i.e., OPA reflections) predicted by the Object-Center Hypothesis preserve the location of the protruding part on the screen.

---

[26] This center is not the same as the center of mass that can be computed for the object.

In my previous study, the procedures for inferring object axes made it possible to probe the tilts but not the origins of the coordinate axes. In each trial images of stimulus objects (i.e., both target and the probe objects) were displayed at the center of the screen. Participants always saw the objects at the center of the screen and were never allowed to move the objects to a different location. As a natural consequence, when an orientation reflection error was made, the center of the reflection axes (i.e., the point where the axes intersected) always corresponded to a fixed location on the object, which coincided with the center of the screen (Figure 4-4B). To probe how the axes' origins are defined, an experiment should be designed so that there is no fixed point on the object that always corresponds to the center of reflections. In other words, there must be several possible locations where a target can appear on the screen, and participants must have the option of placing their response anywhere on the screen. In the next chapter I discuss two experiments (Experiment 3 & 4) for probing the origins of coordinate axes.

CHAPTER 5. INSIGHTS FROM MIRROR-IMAGE REFLECTION ERRORS

5.1. Goals and Motivations

There are two primary goals of this thesis. The first goal is to probe the mechanism for representing the spatial relationships of parts within an object. The first two experiments reported in this chapter (Experiment 1&2) discuss findings related to this goal. These findings suggest that the mechanisms posited in the Coordinate Orientation Representation (COR) theory can be adapted to account for the way object parts' relations are represented in the brain. The second goal of the thesis is to uncover the geometric properties that constrain the way shape coordinate axes are defined. In the latter two experiments of this chapter (Experiment 3&4), I discuss findings relevant to the second goal. The discussion below provides a detailed overview of this chapter.

Probing the Mechanisms for Relating Object Parts

In Chapter 3 of this thesis, I argued that the basic function of coordinate axes is to provide bases for defining spatial directions within an object. To represent relations of parts within an object, the relations must be encoded along multiple independent directions, each direction defined by a coordinate axis. In theories that assume the hierarchical organization of object parts (e.g., Marr & Nishihara, 1978), the way parts' relations are represented is assumed to involve mapping the axes of a local part to the axes of a global part (e.g., the axes of a finger and the axes of the hand). I proposed that the mechanism for mapping the axes could be explained using a framework adapted from a theory in the orientation representation literature. The Coordinate Orientation Representation (COR) theory (McCloskey, 2009; McCloskey et al., 2006) assumes that the brain represents the orientation of a whole object (e.g., the orientation of a teapot in a room) by specifying different parameters for mapping the axes of the overall object (e.g., the principal and the secondary axes of the teapot) to the axes of an extrinsic frame of reference (e.g., the vertical and the horizontal axes defined by the room). Notably, if some encoding failure occurs that affects the parameter for relating the axes' polar features (i.e., the *polarity correspondence* parameter), orientation reflection errors across the object's intrinsic axis are expected (i.e., OPA and OSA reflection errors, see § 3.4.1. Coordinate Orientation Representation (COR) Theory). In previous orientation recall studies, researchers found that errors of this type occurred more frequently than expected by chance (Chaisilprungraung et al., 2019; Gregory et al., 2011; Gregory & McCloskey, 2010).

If the COR mechanism for representing whole-object orientations can be adapted to explain the representation of object parts' orientations, then similar errors involving intrinsic-axis reflections are expected. Specifically, when participants are asked to recall how object parts are spatially related, I expect a number of orientation errors to occur where an object part is reflected across its intrinsic axis (i.e., IPA and ISA errors; see § 3.6.1. Representation of Relative Part Orientation). In Experiment 1, I test this hypothesis and find that the results conform to the prediction: IPA and ISA reflection errors occur more frequently than other types of orientation errors. I interpret the results in light of the adapted COR framework, suggesting that the errors occur because of failures that selectively affect the encoding of relations between the axes' polar features (i.e., *polarity correspondence* parameter). In drawing this conclusion I also take into account an alternative explanation that assumes crude-level shape processing, as well as explanations inspired by computer vision research (i.e., medial axes, convolutional neural networks). In addition to the main result, I also discuss an interesting finding from the experiment which offer important insights about how locations of object parts are defined.

Experiment 2 further investigates whether the adapted COR mechanism can be used to explain the way object parts' locations are represented. In Chapter 3 I propose that the brain represents locations by separately encoding two distance parameters (i.e., *magnitude* and *direction*) for relating axes of object parts (§ 3.6.2. Representation of Relative Part Location). Encoding failures that affect the *magnitude* parameter result in an error where the precise location of a part within an object is incorrectly recalled. Failures that affect the *direction* parameter result in an error where a local object part is positionally-reflected across an axis of a global part. Experiment 2 examines whether the location reflection errors frequently occur when participants are asked to perform a task similar to Experiment 1. The results reveal a systematic occurrence of the location reflection errors, thus supporting the hypothesis that the adapted COR framework can account for the representation of relative part location. Moreover, the experiment presents an unexpected finding which further clarifies the question from Experiment 1, about how object parts' locations are representationally defined.

Uncovering Geometric Basis for Defining Object Axes

The second goal of the thesis is to investigate the geometric basis for determining the overall object axes. Chapter 4 discussed my previous work which demonstrated that the tilts of object axes were determined on the basis of the object's elongated part (Chaisilprungraung et al., 2019). When participants were asked to remember the orientation of an object that was made up of an elongated part and a small protruding part (e.g., a hatchet), reflection errors tended to occur across an axis corresponding to the elongated part (e.g., the handle of a hatchet), rather than the axis that corresponded to the overall elongation of the object in some ways (e.g., axis of least second moment, longest-span axis) (see § 4.2. A Closer Look at The Notion of Elongation). However, in these experiments it was not possible to tell whether the origin of the object axes was also defined based on the center of the elongated part (i.e., Elongated-Center Hypothesis), or the center of the overall object (i.e. Object-Center Hypothesis).

In Experiment 3 and 4, I propose ways of dissociating these hypotheses by modifying the procedures for inferring object axes. Participants see a stimulus object appearing at a random orientation and location on the screen, and are afterwards asked to adjust a probe object by rotating, dragging, and flipping it to put it back in the same orientation and location. Experiment 3 first tests this new procedure on a set of objects with simple elongated shape (e.g., a pen, a knife). If the procedure can be used for inferring the origin of the coordinate axes, the location where the principal and the secondary axes intersect should correspond to the center of the object. Then in Experiment 4, I apply the new procedure on objects that are made up of an elongated part and a protruding part (e.g., a hatchet, a corkscrew), which are also the same objects used in the first experiment of my previous study. The results of Experiment 4 reveal common points of axis intersection that correspond to the center of the overall object rather than the center of the elongated part. I discuss this finding in terms of the theoretical framework proposed in Chapter 4.

5.2. Shape Recall Task

A novel shape recall task is devised to address the questions in Experiment 1 and 2. Participants are asked to remember how parts within an artificial object are spatially related. Participants see three successive views of a stimulus object on the screen in each trial (Figure 5-1). At each view the stimulus appears at a random location on the screen, and the orientation of the whole object with respect to the screen is also selected at random. Importantly, across the three views, the way that the object parts are internally arranged is the same, and participants are instructed to pay attention to this internal arrangement. After the last target disappears participants see a probe stimulus, which is simply the target stimulus with a randomly-selected arrangement of parts. Participants are instructed to adjust the parts of the probe stimulus (i.e., by dragging, rotating or flipping), so that the parts have the same arrangement as in the target. When the answer is an incorrect, the analysis identifies the error pattern.

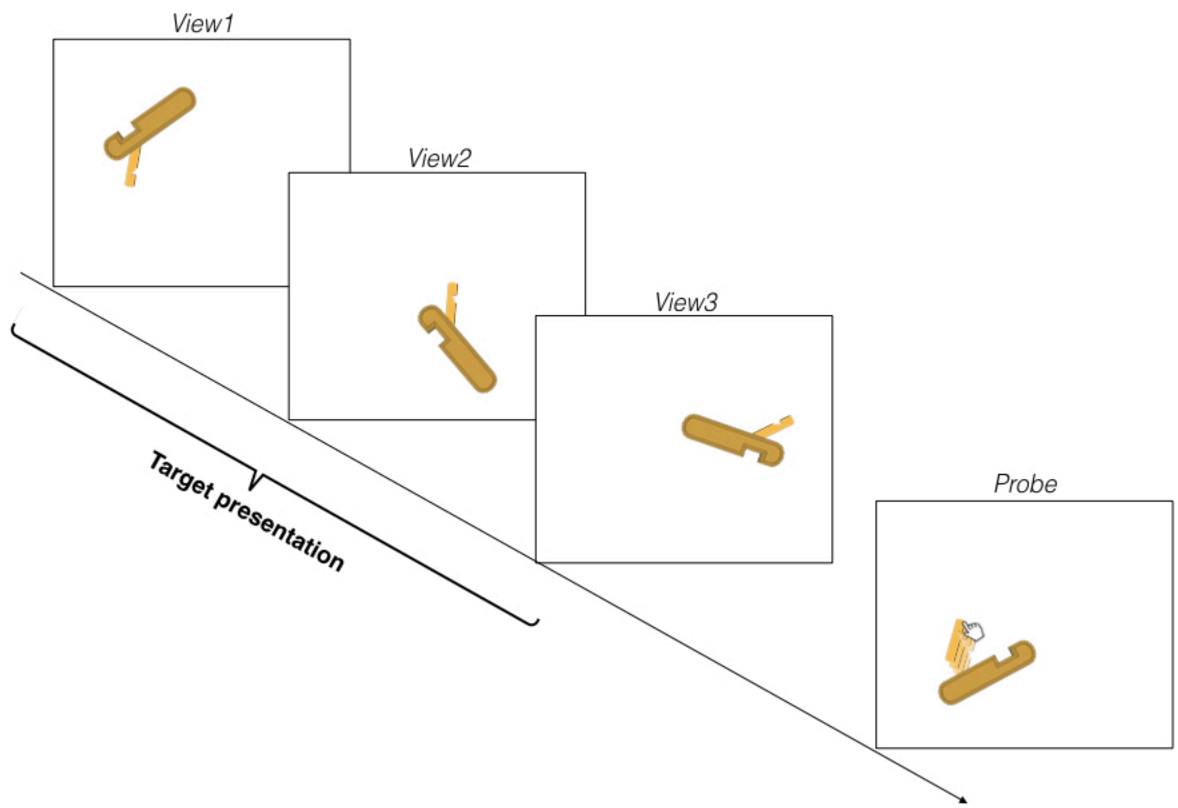

Figure 5-1. *Basic procedure of the shape recall task*

An important feature of the shape recall task is that the target and the probe stimuli are displayed at random locations and orientations on the screen. A critical advantage of this procedure is that participants are forced to remember how object parts are internally related. Since the stimulus' location and orientation change with each presentation, information about where the stimulus appears or how it is tilted relative to the screen (or to any external frame of reference such as the observer's body) is irrelevant to the task. That is, participants cannot succeed in the task by encoding relationships between the stimulus and an extrinsic frame of reference. Therefore, errors in this task can be safely interpreted in terms of failures that occurred when the visual system tries to encode spatial relationships within the object-centered frame of reference.

Participants' responses in the shape recall task are analyzed in two ways. First, the analysis identifies if and what type of orientation error (i.e., error involving how the small object part is oriented relative to the base part) occurs in a trial. Second, the analysis examines whether a location error (i.e., error involving where the small object part appears relative to the base part) occurs in a trial. Experiment 1 focuses on investigating the representations of object parts' orientations, thus the analysis of this experiment takes into account only the orientation data when the object part's location is correctly remembered. Then in Experiment 2, I fully examine the pattern of location errors.

5.3. Experiment 1

5.3.1. Experiment 1 Stimuli and Methods

Participants

Participants were 48 Johns Hopkins University undergraduate students (29 female, 46 right-handed) with normal or corrected-to-normal vision. For all experiments reported in this thesis, participants gave written informed consent in accordance with the procedures approved by the Homewood Institutional Review Board at the Johns Hopkins University. In addition, they received extra credit in a course for participation.

Stimuli

Stimuli, shown in Figure 5-2A, were images of six novel objects designed using the software Google Sketchup (http://sketchup.google.com). Each stimulus object was made up of two parts that could be assembled in different ways to create multiple configurations. The small part of an object could be attached to the side of the base part at six locations, as shown by the stars in Figure 5-2B. At each location, the small part could also have eight different orientations relative to the base (see callout box in Figure 5-2B). These orientations corresponded to tilts that varied by steps of 90˚ (i.e., 45˚, 135˚, 225˚, 315˚ relative to the base's axis of elongation), and their enantiomorphs (i.e., mirror-image reflections).

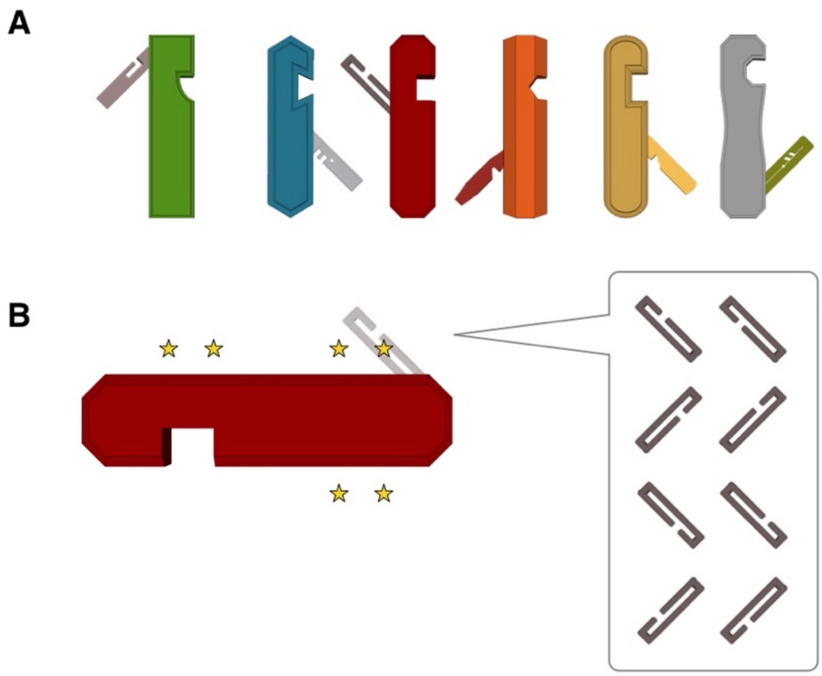

Figure 5-2. *(A) Stimulus objects in Experiment 1. (B) Illustration of possible configurations. For each stimulus object, the small part is attached to the large part at one of six possible relative locations (depicted by the starts). At each location, the small part can have eight possible orientations.*

The combination of six stimulus objects (Figure 5-2A), and 48 part configurations (6 relative locations x 8 relative orientations; Figure 5-2B) created the total of 384 target objects. These targets were counterbalanced and tested across every eight participants, so that each participant saw 48 targets.

Procedure

Each participant completed 24 test trials in a session lasting approximately 30 minutes. On each trial two target objects were presented, each at three different views, and then tested (Figure 5-3). During the target presentation phase, the three different views corresponded to the same target displayed at a random whole-object orientation and at a random place on the screen. The whole-object orientation was sampled randomly (with replacement) from the 720 possible orientations (2 enantiomorphs × 360 tilts). The location of the target was randomly sampled from 16 points in a 5x5 rectangular grid scaled to fit the screen.

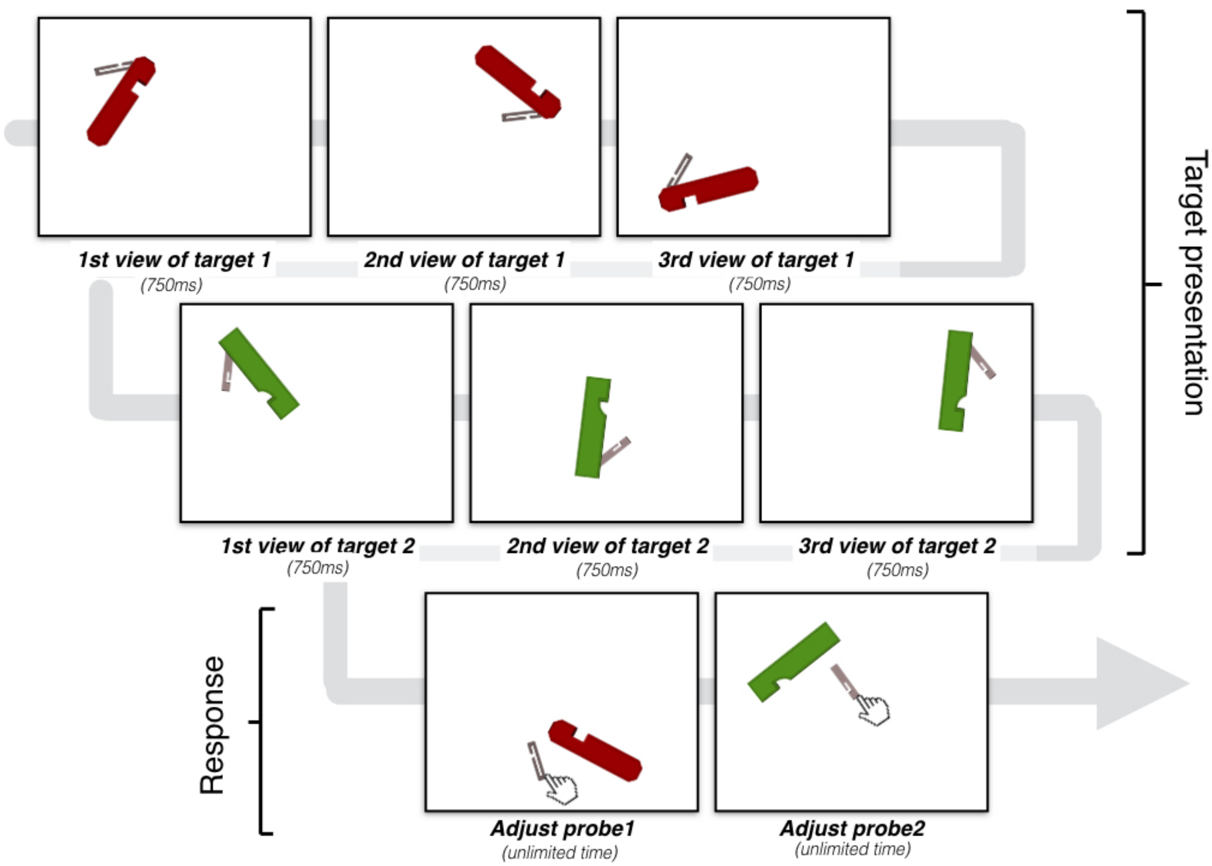

Figure 5-3. *Illustration of a trial structure.*

Immediately following target presentation on each trial, the participant's memory for the target's part configuration was tested. For each target, a stimulus object with the same parts attached in a random way was shown, and the participant modified the part arrangement of this probe stimulus in an effort to match it to the configuration of the previously-seen target. The two targets were tested in the order of their initial presentation (as illustrated in Figure 5-3).

The participant could change the orientation and the location of individual parts in the probe stimulus by clicking on the parts and pressing different keys on the keyboard. They were allowed to change the location of the small part of the probe (but not the base part) by clicking and dragging it to anywhere on the screen. (Figure 5-4A). They could also change the tilt of either part in the probe, by clicking on that part to select it and pressing the right and left arrow keys to rotate it clockwise or counterclockwise, respectively (Figure 5-4B). Finally, they were able to transform any part from one enantiomorph to the other by flipping it. By pressing and holding the shift key and clicking on a part, they flipped the part either lengthwise or sidewise (Figure 5-4C). If a click was made near an end of the part (see top row in the callout box in the figure), the part was flipped lengthwise, or across the axis parallel with the part's width. If a click was made near the part's middle (see bottom row in the callout box), the part was flipped sidewise, or across the axis parallel with the part's length.

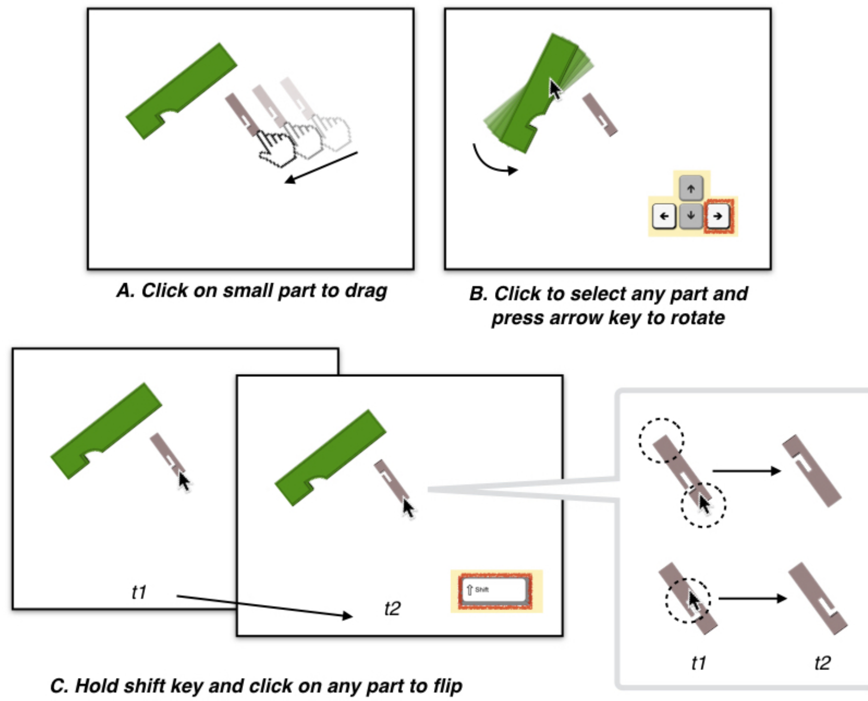

Figure 5-4. *Illustrations of the procedures for modifying the part configuration of the probe. Participants were able to (A) click the small part and drag it to a different location on the screen, (B) rotate any part within an object by clicking the part and pressing the arrow keys, (C) transform a part's enantiomorph by holding down the shift key and click on the part. The drag/rotate/flip procedure did not have to be used in the order shown in the figure in the actual experiment.*

No time limits were imposed on responses, although participants typically responded within approximately 3-4 s of probe presentation. After both responses had been completed, feedback was presented in the form of accuracy scores. The feedback was aimed at encouraging participants to do well in the task, without providing detailed information about how the target and response orientations differed. The maximum score of 100 was awarded for a response showing accurate location and orientation of the parts. Fifty points were allotted to the scoring of orientation accuracy, and another fifty points to the scoring of location accuracy. Responses with the maximum orientation scores were those that had the correct enantiomorph, and whose tilts were within 15° of the correct tilts. For each additional 30° departure from the correct tilt, 10 points were deduced. If a reflection error was made (i.e., a response showing the wrong enantiomorph), another 10 points were deduced. If the point deductions totaled 50 or more, the orientation score was 0. The location score was calculated based on the lengthwise and the widthwise distance from the correct answer. Participants were not informed about how the scores were calculated.

Before the actual experiment participants were given extensive practice, which was designed to help the participants become familiarized with the task, and to ensure that they understood the experiment's instructions. In three practice trials they were instructed to adjust a probe so that it was identical to the target--that is, so that the large parts of the objects had the same orientation, and the small parts had the same orientation. The target remained on the screen while the participants made a response. Then, in two practice trials, they saw three images corresponding to different views of the same target appearing on one side of the screen. Participants were told to modify the probe so that its internal part configuration was the same as the target's configuration--that is, by ignoring the absolute orientation of each individual object part. As before, the images of target objects remained on the screen while the response was made. Finally, the participants saw four practice trials that were similar to the actual experiment. Three images corresponding to different views of the same target briefly appeared on the screen in succession, and participants were instructed to modify the probe to match the previously seen target. In these final practice trials participants started making a response only after the targets disappeared. If a response was incorrect (i.e., if the response scores were below a certain pre-determined threshold), a red box appeared over the screen and the participants had to modify the response until it was correct.

For all studies reported in this thesis, experiments were carried out on a square 19-inch Dell monitor. Stimulus presentation and response collection were controlled by a MATLAB (MATLAB 2014a, The MathWorks, Natick, 2014) script using the PsychToolbox package (Brainard, 1997).

5.3.2. Experiment 1 Results

Participants' responses were first classified according to whether the location of the small object part was correct. For responses with a correct location, further analyses were performed to analyze the pattern of orientation error.

Classifying Location Responses

The plots in Figure 5-5 shows participants' responses across different conditions of target presentation. Each short line represents the relative location and tilt of the small part in an object. The black lines denote the correct answers (i.e., small object parts with a correct location and tilt), and the grey lines denote participants' responses, which were either classified as having either a correct (i.e., dark grey lines) or an incorrect location (i.e., light grey lines).

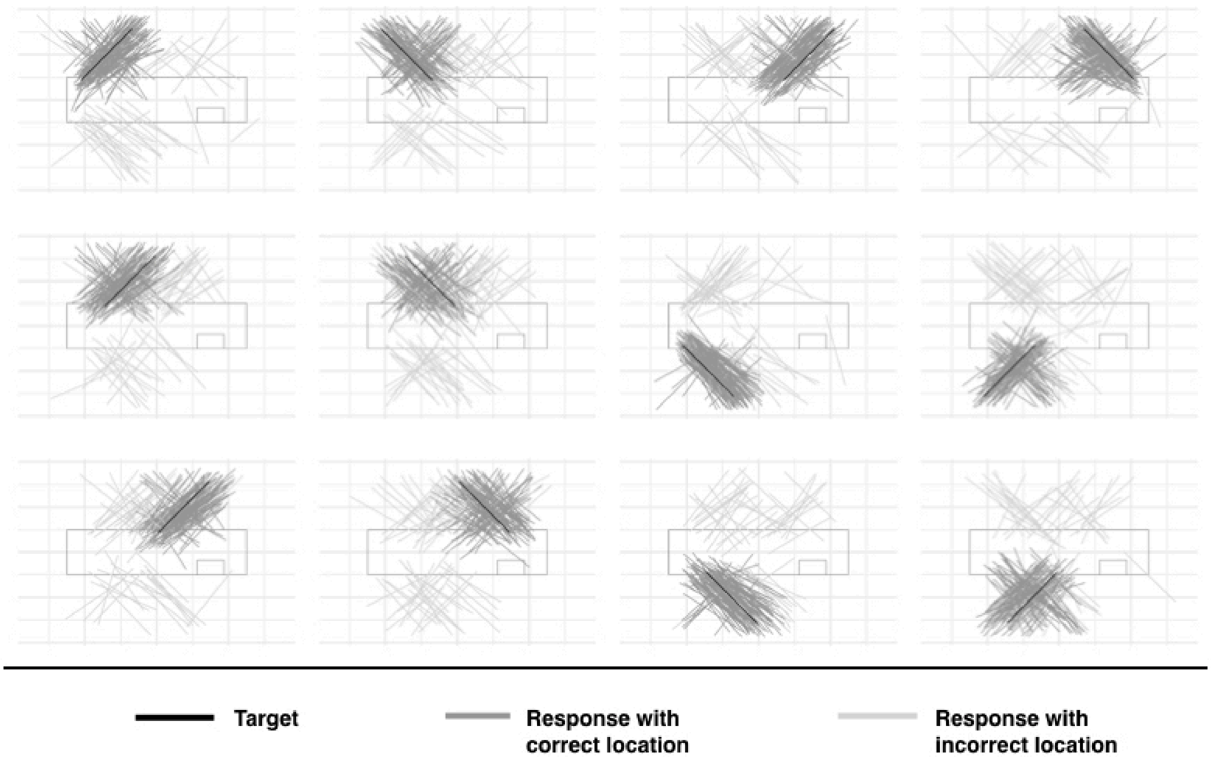

Figure 5-5. *Plots showing the tilt and location of participants' responses. Each line corresponds to each individual response. The responses were classified according to whether the location was correct or incorrect.*

The basis for categorizing the location errors was the absolute displacement of the small object part, which was defined at the part's center. That is, a response was coded as having a correct location if it preserved the location of the target defined at the middle of the small object part. A response whose location (i.e., the location of the small object part's center) was displaced from the target for less than a pre-determined, arbitrarily-chosen threshold of one-sixth the length of the base was classified as having a correct location. A response whose displacement exceeded this threshold was coded as incorrect. This analysis was equivalent to drawing a circle around the target stimulus, and classifying all responses with centers falling within the circle as correct.

Based on this classification, responses with correct locations accounted for 1,593 of the total 2,304 responses (69.1%). Separate ANOVA analyses revealed that the rate of location accuracy did not significantly vary across the six stimulus objects, or across the 48 configurations of objects in the target presentation.

Orientation Distribution for Responses with Correct Locations

When the relative location of the small object part was correct, the orientation pattern was analyzed. To adjust a probe stimulus to create a response, participants were allowed to rotate object parts to whatever degree they liked (see Figure 5-4). Thus, the orientation data was continuous. I first categorized the response orientations according to the nearest possible tilts and enantiomorphs of the target (i.e., see Figure 5-2B for all combinations of tilts and enantiomorphs), and discarded any response whose tilt deviated from the target by more than 30˚, in either direction.[27] The categorized responses were then classified according to the correct answer or the seven possible orientation errors.

Responses with correct orientations accounted for 43.6% of all data with correct location (i.e., 806 of 1,593 responses). For incorrect responses, errors involving reflection of the small object across the intrinsic principal axis (i.e., IPA reflection errors) and the intrinsic secondary axis (i.e., ISA errors) accounted for the highest percentage (i.e., 13.2% and 11.3% respectively; see Figure 5-6). A repeated measures one-way ANOVA revealed a significant main effect of error types ($F(6,329) = 16.23, p = 0$). Post-hoc Tukey HSD comparisons confirmed that both IPA and ISA errors were significantly more frequent than all other errors ($p \leq 0.01$, for all relevant comparisons). The rates of the IPA and the ISA errors did not significantly differ ($p > 0.5$). The

[27] An example of a discarded response is an error where the small part is tilted 90˚ from the base. The discarded errors account for 41 of the total data (2.6%).

predominance of the IPA and the ISA reflection errors suggested that reflection errors across an intrinsic axis of object part was indeed common, thus lending support to the hypothesis that similar mechanisms involving perpendicular coordinate axes are used for representing object parts' relations as for representing the orientations of whole objects. Further analyses revealed that this pattern was consistent across stimulus objects and conditions of presentation (see APPENDIX A.1. Experiment 1).

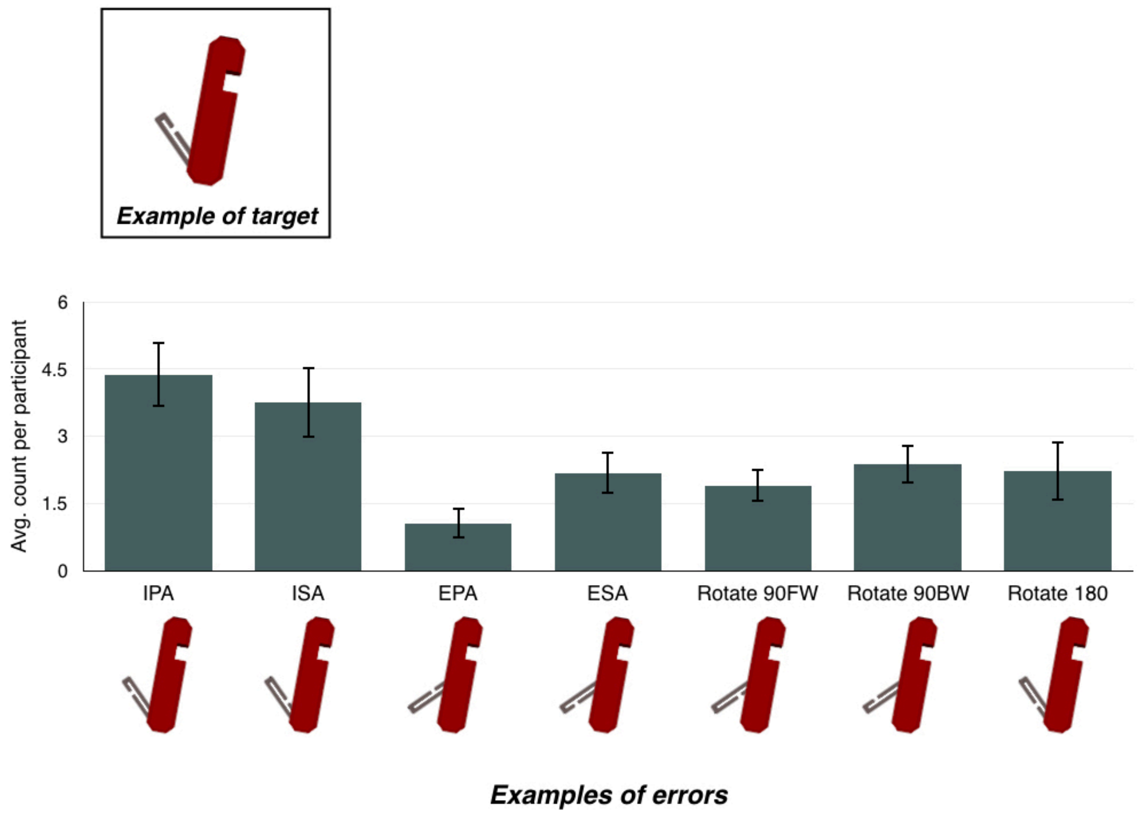

Figure 5-6. *Histogram showing the distribution of part orientation errors, when the relative location of object part was correctly remembered. Error bars indicate 95% CI.*

In addition to the main results, the post-hoc pairwise comparison also revealed that the EPA (Extrinsic Principal Axis) reflection errors, or the errors where the small object part was reflected across the principal axis of the overall object, occurred at a rate significantly lower than other types of errors. The EPA errors were significantly less common than the ESA[28] errors, ($p = 0.07$), and were less frequent than errors involving 90˚ rotation (i.e., 90˚FW & 90˚BW[29]), or 180˚ rotation of the small object part ($p < 0.05$ for all comparisons). No other pairwise comparison was statistically significant.

Orientation Distribution for Responses with Incorrect Locations

Although the main interest of this experiment was orientation errors when locations of object parts were correctly remembered, we also examined orientation errors for responses with incorrect locations. Under the adapted COR theory (§ 3.6), the latter type of responses may be assumed as occurring because of multiple failures affecting orientation and location representations. Compared to responses that occurred as a result of a single representational failure (i.e., IPA and ISA reflections due to failures in encoding polarity correspondence), errors resulting from multiple failures are expected to be less common.

Our analysis reveals a pattern consistent with this prediction. Figure 5-7 shows the orientation distribution of participants' responses when the location was correct and incorrect. We classified a response as having an incorrect location if the distance from the target measured between the centers of object parts was less than a radius which was pre-determined to correspond to one-sixth the length of the base. Compared to responses

---

[28] i.e., Extrinsic Secondary Axis, or errors involving reflection of the small object part across the object secondary axis.

[29] FW and BW are acronyms for forward and backward rotation respectively. These terms signify the directions of rotation that are defined based on the intrinsic structure of the stimulus object. That is, the base part of the object is assumed to have an intrinsic 'front' and a 'back' sides, which are arbitrarily chosen for an object. The reason why traditional definitions of rotation direction, such as clockwise or counterclockwise, are avoided is because these directions must be specified relative to a reference frame extrinsic to the object (e.g., observer-based frame, screen-based frame).

with correct location, the orientation pattern for responses with incorrect location was more evenly distributed.

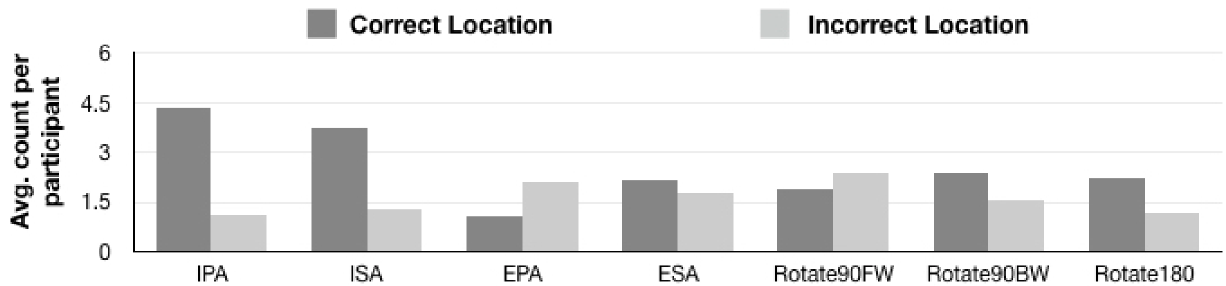

Figure 5-7. *Bar plot of orientation error distribution for responses with correct and incorrect location.*

Alternative Way of Classifying Location Errors

The results reported thus far were obtained based on a certain decision about how object parts' locations are defined. Specifically, a response was classified as having a 'correct location' (or as being a particular type of translation error) if it preserved the location of the small object part defined at the part's center. Figure 5-8A illustrates an example of a response that would be classified as having a 'correct location' under this definition. Here the target and the response coincide exactly at the center of the small object part, and the absolute displacement of the part's location is zero.

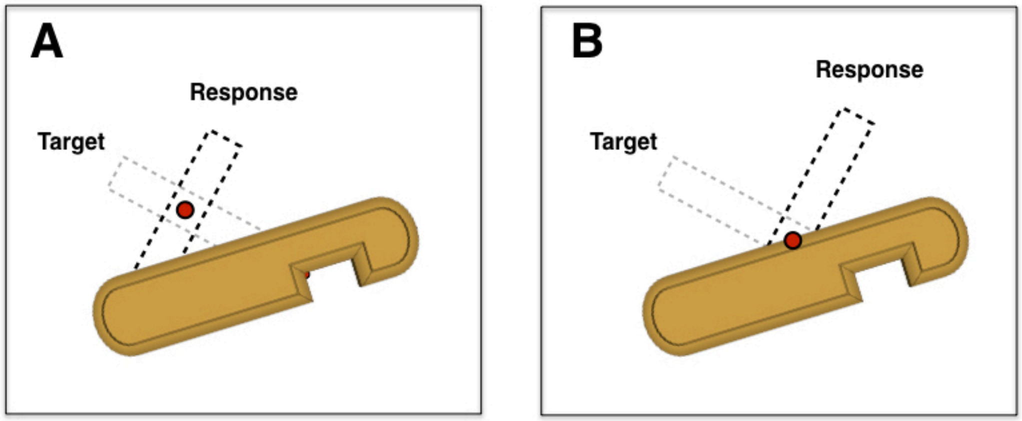

Figure 5-8. *Different ways of defining the location of the small object part.*

An alternative way of defining location errors is on the basis of the place where the parts are attached. That is, a response may be classified as 'correct' if what is being preserved is not the location of the center of the object part, but the place the part is attached to the base. Figure 5-8B illustrates an example of a response that would be classified as 'correct' under the latter definition. Under the first (i.e., the current) definition, the response in Figure 5-8B would be coded as 'incorrect'.

Responses that would be informative in distinguishing the ways of representing object part location were those whose tilts differed from the target. For these incorrect-tilt responses, I compared the average distance between the centers of the object part (henceforth: 'part-center' distance) and the distance between the places of attachment ('part-attachment' distance).[30] If locations are representationally defined at object part's center (Figure 5-8A), the average part-center distance should be significantly less than the part-attachment distance. The reverse pattern is expected if locations are representationally defined at the place of attachment (Figure 5-8B).

[30] The distance referred to in this analysis is the distance across both horizontal and vertical axis.

To conduct this analysis, I first reclassified location data using a new unbiased definition of location errors: Participants' responses were classified as 'correct' if *either* the part-center distance or the part-attachment distance was less than the pre-determined threshold.[31] Figure 5-9 shows the results of this new classification. Responses with correct locations are here shown in either blue or red. Blue responses correspond to responses whose tilts are parallel with the target (i.e., shown in black).[32] Red responses correspond to those whose tilts are opposite to the target (i.e., incorrect tilt).[33] Based on preliminary inspections of Figure 5-9, responses with incorrect tilts (i.e., red) appear to preserve the location at the center of the object part.

[31] i.e., one-sixth of the length of the base.
[32] These responses are the IPA error, the ISA error, the 180˚ rotation error, and the correct response.
[33] These responses are the EPA error, the ESA error, and the 90˚FW & 90˚ BW rotation error.

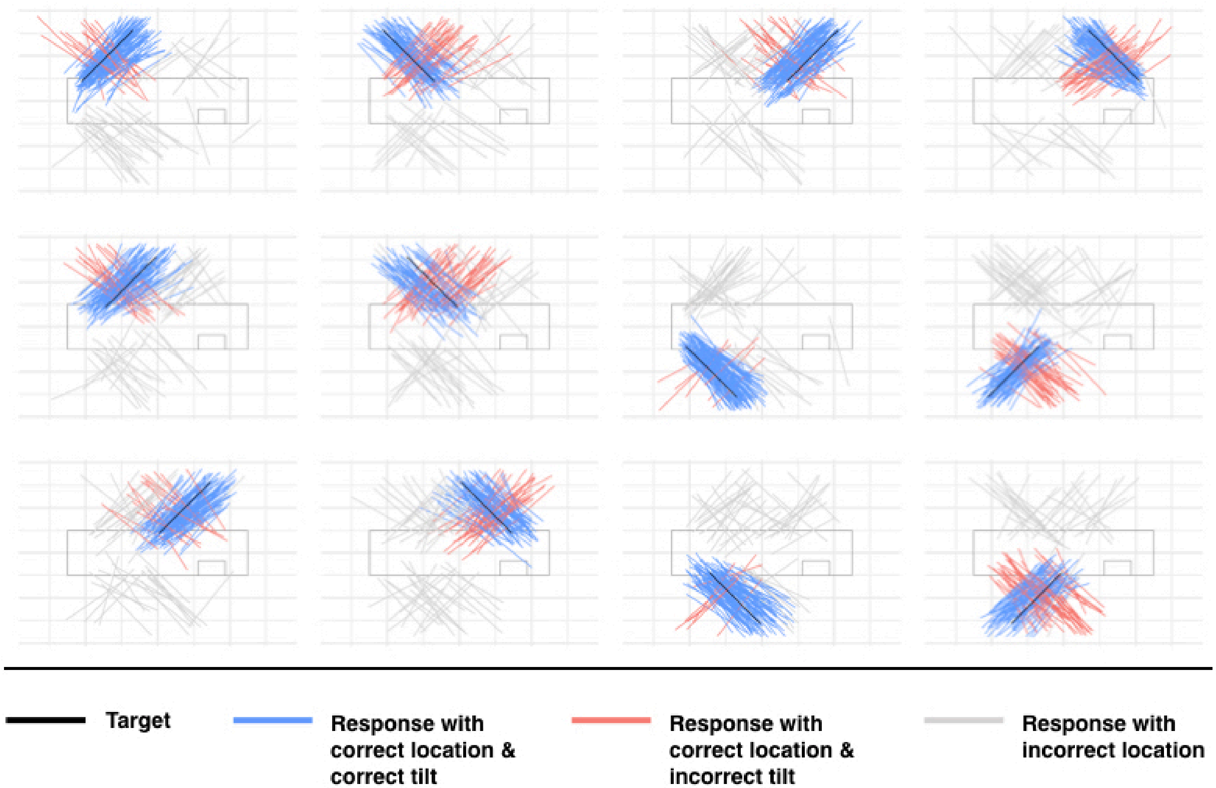

Figure 5-9. *Plots illustrating responses with correct locations. The responses had either correct (blue) or incorrect (red) tilt.*

Statistical analyses verified this pattern. When participants' responses had a correct location but incorrect tilt (i.e., red), the average part-center distance was 0.12[34], which was approximately half of the part-attachment distance ($\bar{x}$=0.23) ($t(2857) = -9.1; p < 0.001$). In short, the results suggested that what was being preserved in participants' responses was the location defined at the center of the small object part, rather than the place at which the parts are attached. The finding also provided a firm grounding for our use of object part's center as a basis for classifying the location data.

[34] The unit of the measurement is the length of the base. That is, the distance of 1 unit is equal to the length of the large object part.

5.3.3. Experiment 1 Discussion

Experiment 1 asked whether the mechanism for representing object parts' relationships is similar to that used for representing orientations of whole objects. We obtained evidence in favor of this view. In particular, errors involving a reflection across an intrinsic axis of object part (i.e., IPA and ISA errors) were observed at high frequency. This pattern was consistent with the findings of the previous orientation recall studies where participants' errors often involved a reflection across an intrinsic axis of the whole object (i.e., OPA and OSA errors; Chaisilprungraung et al., 2019; Gregory et al., 2011; Gregory & McCloskey, 2010). More broadly, the results suggested that the COR theory could be adapted to account for how relative orientations of object parts are represented. The way that the brain encodes the relative orientation of the small object part to the base part may involve assigning a set of perpendicular axes to each part, and encoding the relations between the parts' axes via different COR parameters (see 3.6.1. Representation of Relative Part Orientation). An erroneous encoding of the parameter for mapping the axes' polar features (i.e., *polarity correspondence*) leads to an error where the small object part is reflected across its intrinsic axis, the type of error which is being observed in the current experiment.

Experiment 1 yielded a clear pattern of results and demonstrated that the novel shape recall task could be valuable in studying the nature of object shape representation. Below, I address plausible alternative interpretations of the data, and discuss the remaining questions concerning other interesting aspects of the findings.

Alternative Explanation of the Current Data

An important question that can be asked is whether the current results could be explained by alternative theories that do not assume the notion of the coordinate shape axes. A simpler explanation for the IPA and the ISA errors, for example, may be that the participants remembered only the crude overall shape of the visual stimuli. As depicted Figure 5-10 below, the IPA and the ISA errors have a very similar crude form to the target stimulus. If what participants had remembered from each trial's target presentation was some crude global features of the object as shown in Figure 5-10A, then it was possible that the high rates of the IPA and the ISA errors were simply due to the coarse-level visual similarity of the stimuli.

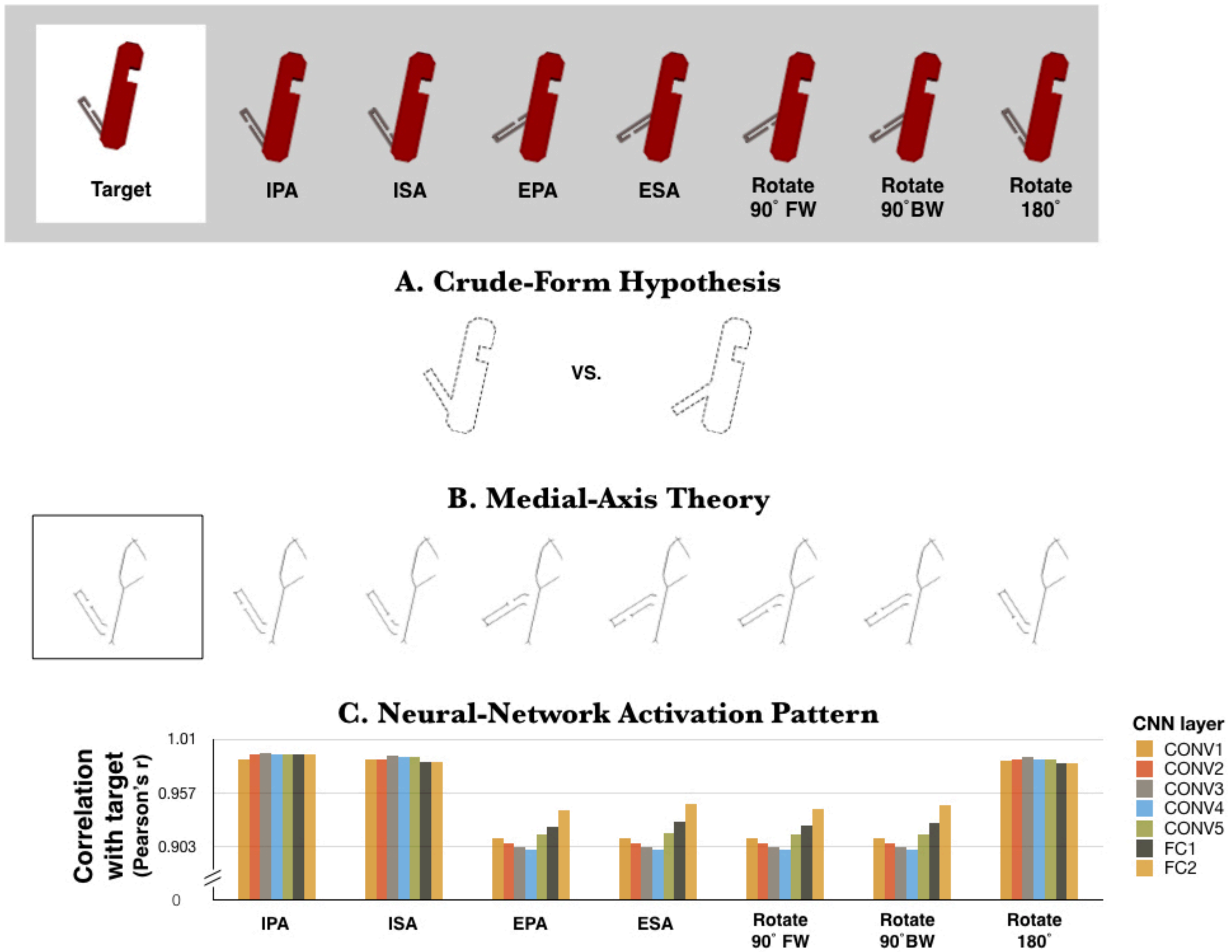

Figure 5-10. *(A) Illustration of a crude form of a stimulus object. (B) The stimulus objects with the corresponding medial structures. The medial axes for these objects were computed using the Average Outward Flux (AOF) skeletons (https://github.com/mrezanejad/AOFSkeletons) (Rezanejad & Siddiqi, 2013). (C) Bar plot showing pairwise correlations of activation patterns across different layers of CNN. Error bars denote 95% CI.*

However, some aspects of the current data could not be accounted for by the crude-form hypothesis. According to this hypothesis, errors that involve 180˚ rotation of the object part (i.e., 180˚ rotation error) are predicted to occur at a rate comparable to those of the IPA and the ISA, as these errors have equally similar crude form to the target's (see Figure 5-10A). As our data indicated, however, the 180˚ rotation errors occurred at a significantly lower rate than the IPA and ISA errors (see Figure 5-6 in the Results section). Indeed, the rate of the 180˚ rotation errors did not significantly differ from the rates of other errors with a different global form (i.e., 90˚ rotation errors, ESA errors). In light of these results, the crude form hypothesis provides an unlikely explanation for our finding.

Another potential explanation is a theory that assumes the notion of medial axes. As discussed in chapter 3, some theorists posit that the brain represents object shapes in terms of axes computed from shape contours (e.g., Ayzenberg et al., 2019; Blum & Nagel, 1978; Firestone & Scholl, 2014; see § 3.3). Due to extensive research in computer vision on medial axes, there exist different algorithms that may be used to compute medial axes for a shape, each yielding a somewhat different version of medial axes (Feldman & Singh, 2006; Geiger et al., 2003; Katz & Pizer, 2003; Kimia, 2003; Ogniewicz & Kübler, 1995; Rezanejad et al., 2019; Siddiqi & Pizer, 2008). To illustrate a few examples of the axes for our stimuli, I used an algorithm that was recently developed based on the mathematical notion called 'Average Outward Flux' (i.e., AOF skeletons; Rezanejad et al., 2019). The AOF axes have also been demonstrated be useful in scene classifications (Wilder et al., 2019).

Figure 5-10B shows the AOF medial axes for each stimulus. An important problem faced in this analysis is that it is not clear how the similarity between skeletal axes should be quantitatively measured. Perhaps a simplistic but reasonable way of measuring the axes' similarity is to compute the degree of the visual overlap. That is, similarity may be computed on based on how much overlap exists between medial axes on a pixel-by-pixel basis. However, this method is unlikely to produce results that substantially differ from the crude-form predictions. For the 180˚-rotation error, the pixel-by-pixel overlap is higher when compared with the target and the errors that share a similar crude form to the target (i.e., IPA, ISA, 180˚), than with errors with a different crude form (i.e., EPA, ESA, 90˚FW, 90˚BW). Thus, in the absence of a method for computing medial-axis similarities that produces a result conforming to currently observed error pattern, it is not clear how the Medial-Axis theory can account for the present finding.

Finally, recent advances in computer vision offered a way of quantifying visual similarity across image stimuli. Convolutional Neural Networks (CNNs) are a class of machine learning algorithm that is highly effective in areas of object recognition and classifications. A CNN model can take an image as an input, and through multiple layers involving basic operations (e.g., image convolution, pooling, linear-nonlinear transformation), extract features of the image useful for classification. Each layer of CNN is composed of a set of activation units (i.e., the so-called 'neurons'), which supposedly carry some meaningful representation about the image. Might it be possible that at some level(s) of network processing, CNNs compute crude features of object shapes relevant for solving the shape recall task (i.e., if indeed crude features were assumed to be involved in the task)?

To address this question I analyzed the CNN activation patterns to different pairs of artificial stimulus objects. Alex Net (Krizhevsky et al., 2012) is a widely-used CNN that was trained on more than a million images from an online database[35], and is capable

[35] ImageNet (image-net.org).

of classifying visual inputs into 1000 object categories (e.g., pencil, mouse, toilet paper, many breeds and types of animals) with high accuracy[36]. I tested the pre-trained Alex Net on the artificial objects, and computed the similarity of the activation patterns for every possible pair of objects that corresponded to the target and the response (see Appendix B for full details of the analysis). AlexNet contains five convolution layers (i.e., CONV, see Figure B- ii), and three fully connected layers (FC), with the last FC layer containing units corresponding to category labels. I extracted the activation patterns from each layer except the last layer and computed the pairwise correlations for each layer. The result, shown in Figure 5-10C, revealed a strikingly consistent pattern of activation across CNN layers: Stimuli corresponding to the target were treated as being highly and equally similar to stimuli corresponding to IPA, ISA, and 180˚-rotation errors (average Pearson's r = 0.989, 0.994, 0.990), and these similarities were greater than those of stimuli corresponding to EPA, ESA, 90˚FW, and 90˚BW errors (average Pearson's r = 0.914, 0.916, 0.914, 0.915).[37] In other words, the CNN's responses were contrary to the error pattern that participants produced in the shape recall task, but were consistent with the prediction of the crude-form hypothesis. The results therefore suggested that it was not possible to explain the current data using crude shape features that may be encoded in sophisticated computer vision algorithms like CNNs. Rather, the current finding suggests that high-level shape representations are involved in the shape recall task.

In short, the current experiment provides empirical support for the view that the brain represents object parts' orientations based on systems of coordinate axes. The

[36] AlexNet outperformed all other competitors by a wide margin at an image classification contest in 2012 (ILSVRC-2012).

[37] Further data inspections revealed that this pattern was highly robust across types of object and conditions of target presentation (see Figure B- iii in APPENDIX B.1. Experiment 1).

discussion in this section suggested that the current results could not be sufficiently explained by the Medial Axis Theory, or views that assume crude shape similarity—whether such similarity is computed using a simple means or a sophisticated computer algorithms. Next, I discuss remaining questions of the study.

Remaining Questions

In addition to the main finding, the current experiment presented an interesting result worth further discussing. Specifically, when participants made an orientation error, their error tended to preserve the location of the object part defined not at the region where the part was connected to the base, but at the part's center (Figure 5-9). The finding suggested that the location of the object part was representationally defined at the center of the part. In the model for location representation I proposed in Chapter 3 (see § 3.6.2. Representation of Relative Part Location), the results also have important theoretical implications. In particular, the way that the brain represents the location of a local object part may be by relating the coordinate origin of the local object part, which is defined at the *part's center,* to the coordinate origin of the global object part.

An interesting question that may be raised based on this finding is: Might the pattern persist if the stimulus objects are modified in some way so that the region of attachment is visually more salient? For instance, will the participants continue to represent the location of object part at the part's center, if a visible joint is added to the region where the small part is attached to the base? In Experiment 2, I modified the object stimuli to test this possibility.

Another general question of the study has to do with what can be learned about the mechanism for representing object parts' locations, based on the errors from the shape recall task. In the model of location representation mentioned above, the location is represented by separately encoding two parameters (i.e., *magnitude* and *distance*) for relating the distance between the origins of the coordinate axes. If some failures occur in the encoding which affects the *direction* parameter, then an error involving a location reflection of the small object part across an axis of the base part is expected (see § 3.6.2).

For the current experiment, there were some aspects of the object stimuli that made the experiment less than ideal for probing the location errors. Particularly, the base part of the object possessed a large notch near one end (see Figure 5-2). The reason why we included the notch in the first place was because it was necessary to make the base part bilaterally asymmetric.[38] However, the presence of the notch meant that some types of location errors were obviously implausible. Specifically, participants were unlikely to place the small object part at the location near or coincided with the notch, because they knew that such a response would be incorrect. This constraint on the response location ruled out the possibility of some types of reflection errors. If, for instance, the small object part was presented on the same side as the notch lengthwise, but on the opposite side widthwise, an error in the form of a reflection across the base's principal axis was unlikely. In the next experiment, the object stimuli were modified to remove the constraint.

5.3. Experiment 2

Experiment 2 probed how the relative locations of object parts are represented using the test stimuli modified from Experiment 1. The stimuli of the current experiment no longer had a large notch near an end of the base (see Figure 5-12 for the complete

[38] i.e., asymmetric across both lengthwise and widthwise directions.

stimulus set). Instead, we included a notch at the middle of the object part, and added a distinctive structural feature (e.g., a curve, a prong-like structure) at an end of base to distinguish it from the other end. The stimuli of this experiment were thus bilaterally asymmetric, while also lacking the constraint regarding where the small object part could appear relative to the base.[39] Therefore, several types of location reflection errors were possible.

If a location reflection takes place across the secondary axis of the base part (Figure 5-11, left), then we expect an error where the small object part appears on the wrong side of the base, along the lengthwise direction. Importantly, for the responses that appear on the wrong side, the distance measured from the middle of the base part (i.e., the distance '*x*' in the figure) should be the same as the distance to middle of the target. Similarly, if a location reflection takes place across the base's principal axis, the response that appears on the wrong side along the widthwise direction should preserve the distance from the base's middle. Put it simply, when the target is presented near the middle of the base, the reflected response should also appear near the middle. And if the target is presented at a far location, the reflected response should likewise be far.

[39] That is, except for the central location of the base where the notch was present.

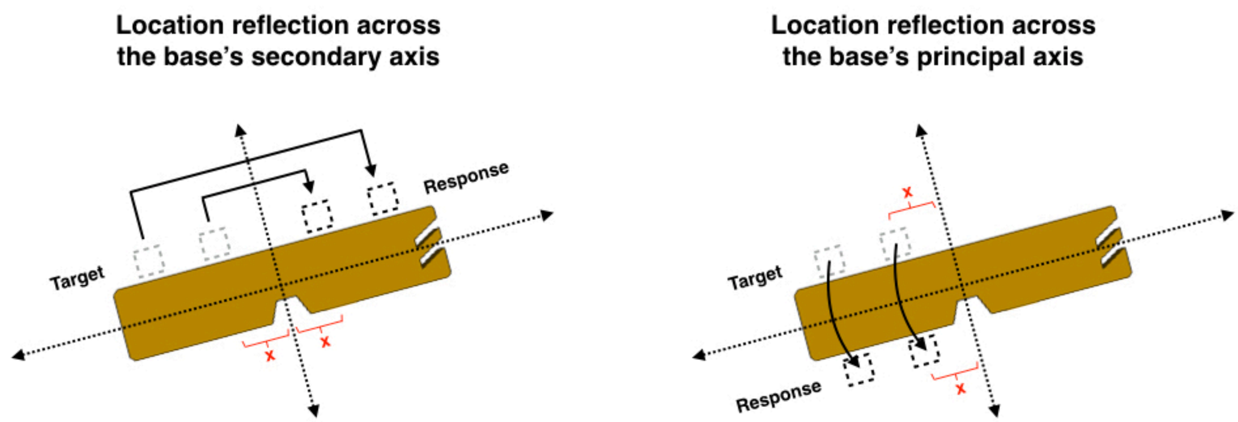

Figure 5-11. *Predictions for location reflection errors in Experiment 2.*

Besides the main goal, the experiment also investigated whether having a salient feature at the region where the parts were attached altered the way the object part's location was representationally defined. Stimulus objects in this experiment had a visible joint at the place of attachment. If the presence of the joint affects the way locations of object parts are represented in any way, we expect participants' responses to no longer preserve the location defined at the object part's center.

5.3.1. Experiment 2 Stimuli and Methods

Unless explicitly noted, the method of Experiment 2 was the same as that of Experiment 1.

Participants

Participants were 60 Johns Hopkins University undergraduate students (46 female; 56 right-handed) with normal or corrected-to-normal vision.

Stimuli

Stimuli were modified from the previous experiment in two ways. First, instead of presenting one large notch near an end of the large object part, we presented the notch at the center of the base part (Figure 5-12). To keep the base part bilaterally asymmetric, we added some structural features (e.g., a curve, jagged tip) to one end of the base, in order to distinguish it from the other end. In addition, we added a semi-circle joint at the place where the small object part was attached to the base. In the actual experiment, the joint appeared only when the object parts were seen as attached, and not when they were detached (as when the participants moved the parts around to make a response).

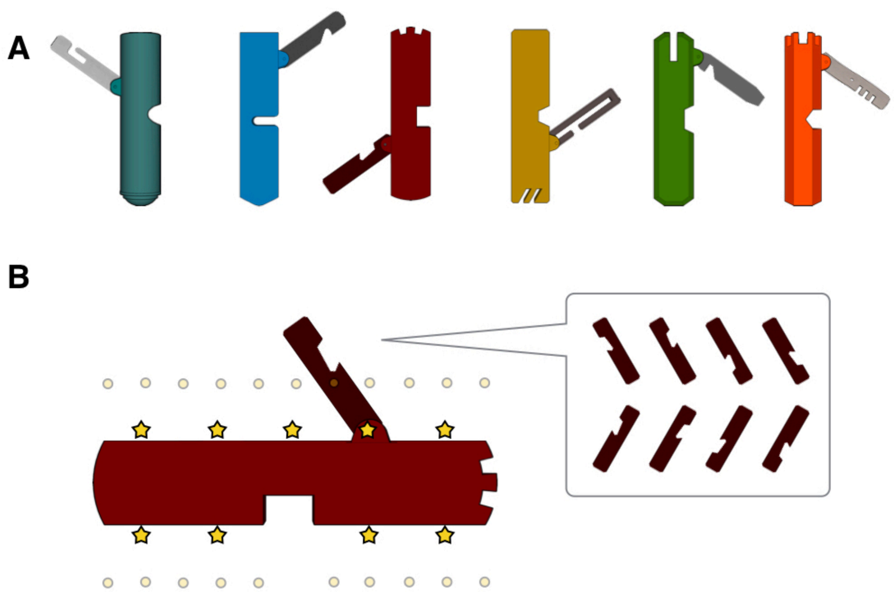

Figure 5-12. *Stimuli of Experiment 2*

The second change in the stimulus design had to do with where the small object part appeared in a target. Contrary to the previous experiment, the possible locations of the small object part (defined either at the part's attachment or at the part's center) were selected so that they were evenly spaced across the two sides of the base. On the flat side of the base (i.e., the side without the center notch), there were five possible locations of where the small part could be attached to the base (see yellow stars in Figure 5-12B), and eleven possible locations that corresponded to the small part's center (see semi-transparent circles). On the side with the center notch, there were four possible locations for the place of attachment, and ten possible locations for the part's center.

The small object part was always tilted at an angle 55˚ relative to the base. This angle was specifically chosen because it allowed *both* types of possible locations (see the stars and circles in Figure 5-12B) to be evenly spaced. For each presentation location defined at the part's attachment, there were eight possible orientations of the small object part (see inset box in Figure 5-12B). The combination of six stimulus objects, nine presentation locations (defined at the part's attachment), and eight orientations resulted in 432 stimulus combinations. These combinations were counterbalanced and presented across every twelve participants, such that each participant saw 36 combinations.

Procedure

Testing procedures were mostly similar to Experiment 1, except for three modifications that were made to improve the methods for adjusting the probe. In the previous experiment participants were allowed to manipulate object parts individually. That is, by using the mouse and the keyboard, they could drag, rotate, or flip an individual part of an object. However, even when the object parts were right next to one another, such that they appeared as if they were attached, participants were unable to manipulate the parts to make them move or rotate together in tandem. This procedural limitation was contrary to the way parts in real-world objects behave.

The current experiment modified the response procedure to allow for holistic transformations of the object. When the small object part was placed somewhere on the screen near the large object part (see Figure 5-13A), the parts would appear as attached via a semi-circle joint. Any adjustment made to an individual part while the parts were attached (i.e., rotate or flip) resulted in whole-object transformations (see Figure 5-13B&C, right panels).

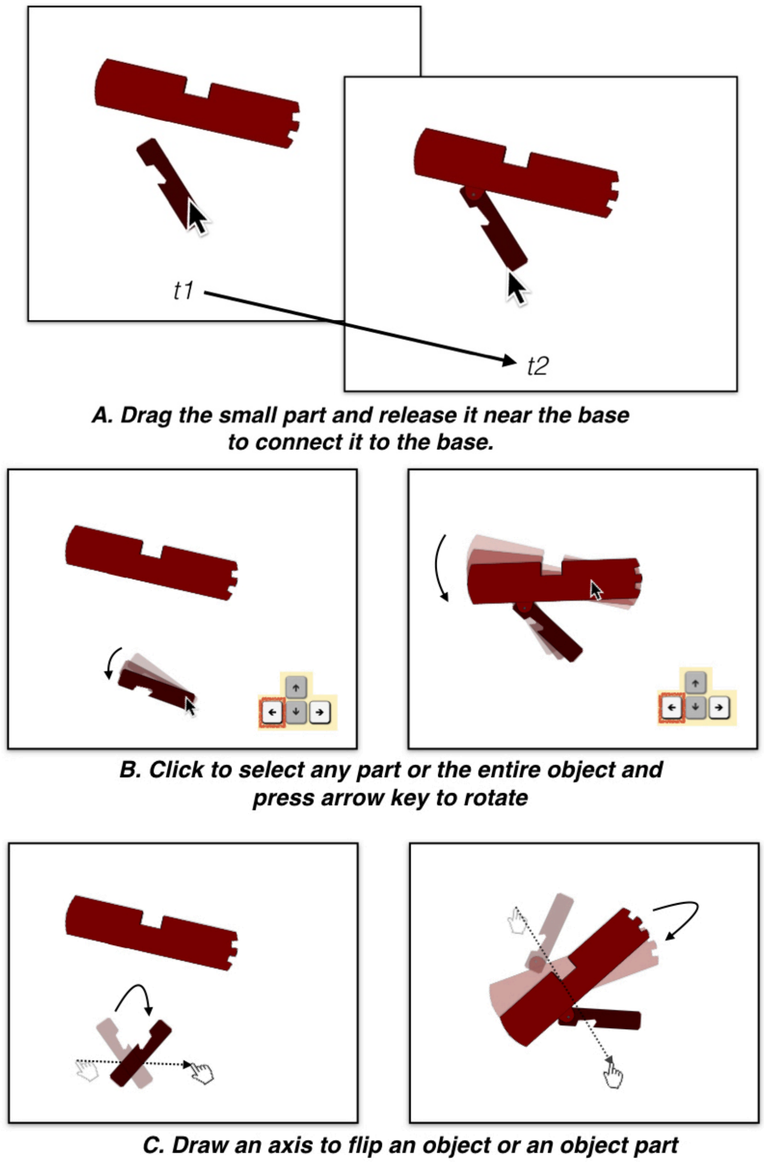

Figure 5-13. *Procedures for adjusting probe stimulus in Experiment 2*

Another important change in the response procedure was the way the stimulus was flipped. In the previous experiment, participants flipped an object part by clicking on the part while pressing the shift key (see Figure 5-4). A click near an end of the part resulted in a lengthwise flip—or a flip a across the axis parallel with the part's width. A click near the middle resulted in a sidewise flip—or a flip across the axis parallel with the part's length. In this original procedure, participants had very limited options regarding how they wanted to flip a stimulus.

The current experiment allowed the participants to 'draw' an axis indicating how the stimulus was to be flipped. Specifically, participants right-clicked twice on the screen to mark the beginning and the end point of an axis (see Figure 5-13C). The stimulus was flipped across the axis orthogonal to the axis that the participants drew.[40] Participants could flip the object parts individually when they were detached (see the left panel of the figure), or as a whole when they were attached (see the right panel).

5.3.2. Experiment 2 Results

To analyze whether a location reflection error occurs in a trial, it is necessary to first decide how we will define the location of an object part. To do this, we asked whether the participants representationally defined the location at the center of the object part, or at the place where the parts were attached. An analysis similar to Experiment 1 was conducted where we compared the part-center distance and the part-attachment distance of participants' responses. The results of this analysis formed a basis for the way location data was analyzed later on in the experiment.

How are Object Parts' Locations Representationally Defined?

Similar to Experiment 1, participants' responses were first classified according to whether the location was correct or incorrect using an unbiased definition: A response

[40] The reason why we had the participants draw the axis orthogonal to the flipping axis (as opposed to having them draw the flipping axis directly) was because this procedure felt more natural. Participants were told they could also think of this procedure as if they were to 'swipe' a shape on a tablet device (e.g., iPad) to flip it.

was coded as correct if either the part-center distance or the part-attachment distance was less than the threshold of one-sixth the length of the base (see discussion in Experiment 1 Results). Figure 5-14 shows the results of the classification. The responses with correct locations (displayed in either blue or red) were further grouped according to whether the tilts were correct (i.e., blue) or incorrect (i.e., red).[41]

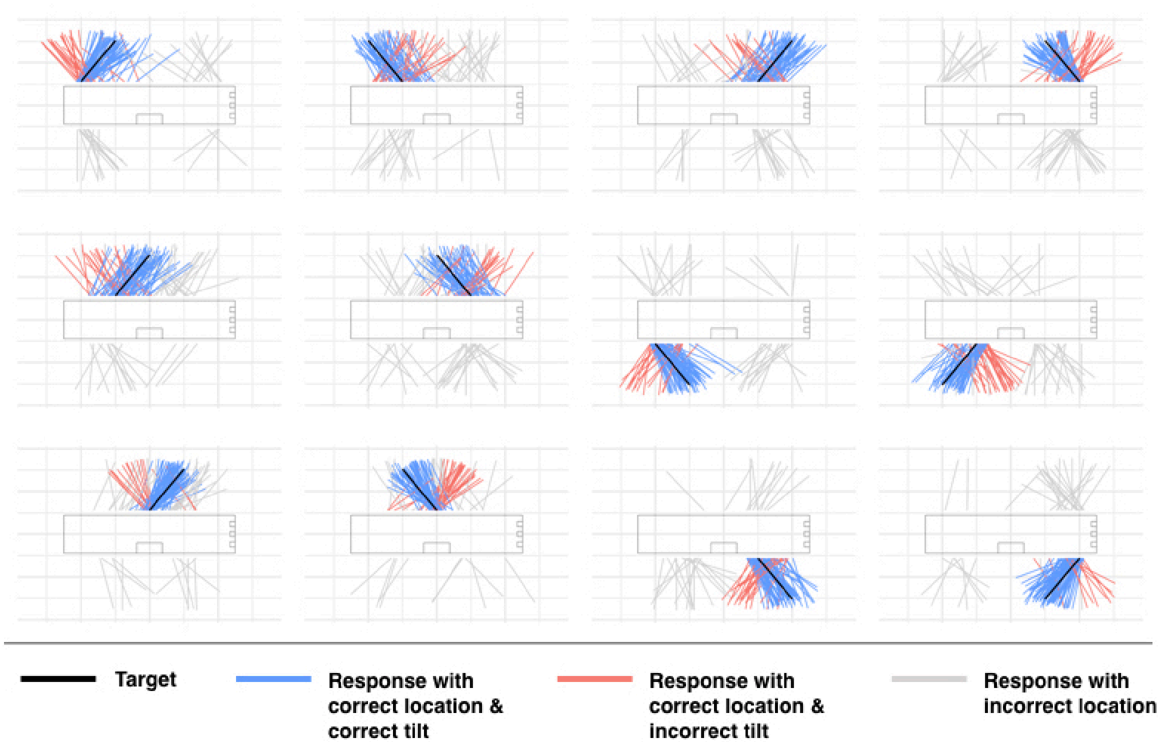

Figure 5-14. *Plots illustrating responses with correct locations. The responses had either correct (blue) or incorrect (red) tilt.*

Based on a preliminary inspection of the results in Figure 5-14, participants' responses appeared to preserve the location defined by the place of attachment rather than

[41] It should be noted that Figure 5-14 displays the results for only 12 target conditions, whereas the actual experiment had 18 target conditions (i.e., 9 presentation locations x 2 tilts, see Figure 5-12). Four conditions were excluded from this analysis because it was not possible for the participants to make a response that preserved the object part's center. An example of the excluded conditions was when the small object part was attached at an extreme end of the base part with its tip pointing outward.

the object part's center. Indeed, when the tilt of the object part was incorrectly remembered (i.e., red responses), the part-attachment distance ($\bar{x} = 0.076$) was approximately half of the part-center distance ($\bar{x} = 0.14;\ t(502) = -9.72, p < 0.001$).

The current analysis revealed a pattern different to what was observed in the first experiment: When a visible joint was added to the region where the object parts were connected, the location of object part appeared to be representationally defined at the place of attachment, rather than the object part's center. The difference in the pattern might be due to the addition of the joint at the point of attachment. On the basis of this finding, we selected the place of attachment as a basis for classifying the location data.

Location Reflection Errors

We analyzed the pattern of location reflection errors. Participants' data were first grouped according to whether the small object part was presented at the far or the near location in the stimulus (Figure 5-15). A target location was counted as 'far' if the place of attachment had a distance of 0.4[42] from the middle of the base part (see the top panel of the figure), and 'near' if the distance from middle was 0.2 (see the bottom panel). Based on this near vs. far grouping, we further classified participants' responses and computed the average distances to the middle of the base part.

[42] The distance's unit is relative to the base's length.

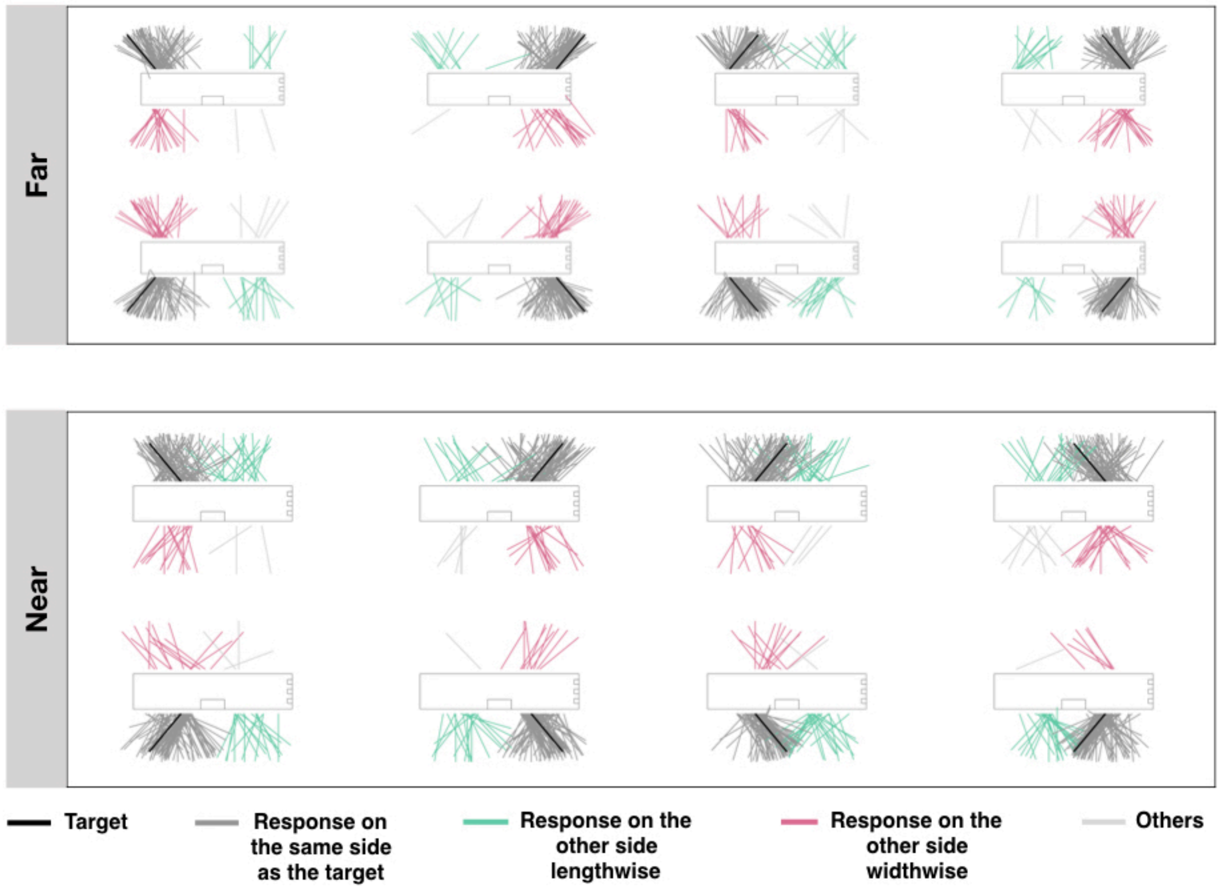

Figure 5-15. *Plots showing participants' responses in Experiment 2. The responses were grouped according to whether the target's location (defined at the place of attachment) was far or near the middle of the base, and then according to whether the responses appeared on the same or opposite side of the target.*

Participants' responses were classified according to whether the location of the small object part, defined at place where the parts were attached, was on the same (i.e., dark grey lines) or different (i.e., green and pink lines) sides as the target. We divided the base part at middle into four quadrants. The responses where the place of attachment appeared in the same quadrant as the target were classified as being on the same side. The responses where the place of attachment appeared in the two adjacent quadrants were classified as being on the other side, either along the lengthwise (i.e., green lines) or the widthwise (i.e., pink lines) directions of the base.

For each type of response, we computed the average distance of the places of attachment and the middle of the base. Figure 5-16 shows a bar plot summary of the results. When the target was presented at the far location (i.e., at 0.4 from the base's middle), the responses also tended to be far, regardless of whether they were on the same or the other side as the target ($\bar{x}$ =0.36, 0.35 and 0.32 for the grey, pink and green responses, respectively). When the target was presented at the near location (i.e., 0.2 from the base's middle), the responses also tended to be near ($\bar{x}$ =0.2, 0.2 and 0.18 for the grey, pink and green responses). Across the near vs. far conditions, there were significant differences in the average distances from middle (i.e., $t(11.6) = -28.2; p < 0.001$ for the grey responses; $t(9.7) = -8.6; p < 0.001$ for the pink responses; and $t(13.5) = -9.7; p < 0.001$ for the green responses). In short, the pattern indicated that location reflection errors frequently occurred.

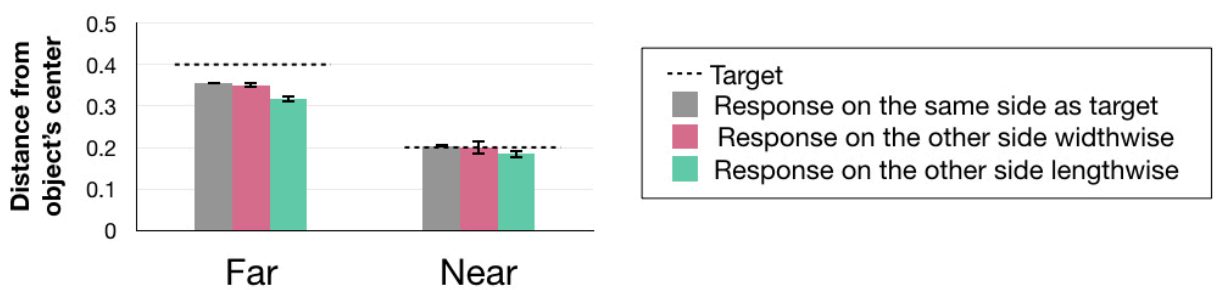

Figure 5-16. *Plots showing the average distances from the middle of the base part for different types of responses.*

In addition to this main finding, we also found a tendency of the participants to undershoot the target's location. When the target was presented at the far location, the responses that were reflected along the widthwise direction of the base (i.e., pink responses) tended to appear at a location nearer where they should have been ($t(7) = -7.1; p < 0.001$). The same was true for the responses that were reflected along the lengthwise direction ($t(7) = -8.6; p < 0.001$). However, we also observed the same tendency to undershoot when no reflection error was made (i.e., grey responses; $\bar{x} = 0.36; t(7) = -16.0; p < 0.001$). These patterns therefore suggested that the tendency to undershoot was not specific to location reflection error. Rather, the results were more likely due to a general tendency of participants to avoid placing the small object part too close to the tip of the base.

Orientation Distribution for Responses with Correct Locations

We examined whether the orientation distribution for correct-location responses was similar to the previous experiment. We first examined whether the locations of participants' responses were correct. Unlike Experiment 1, a response was counted as correct if the part-attachment distance (rather than the part-center distance) was less than the pre-determined threshold corresponding to one-sixth the length of the base. Based on this classification, the responses with correct locations accounted for 71.6% of all data (i.e., 1,547 of 2,160 responses).

Figure 5-17 displays the distribution of orientation errors for when the location was correct. Similar to Experiment 1, we found that errors involving a reflection across an intrinsic axis of the small object part (i.e., the IPA and ISA errors) occurred at high rates (i.e., 10.7% and 14.2% for IPA and ISA respectively). A repeated measures one-way ANOVA with post-hoc Tukey HSD comparisons revealed that the IPA and ISA errors occurred more frequently than most other errors, including the 180˚-rotation errors ($p \leq 0.01$, for all relevant comparisons). The rates of the IPA and the ISA errors compared to one another differed at a marginally significant level ($p = 0.05$).

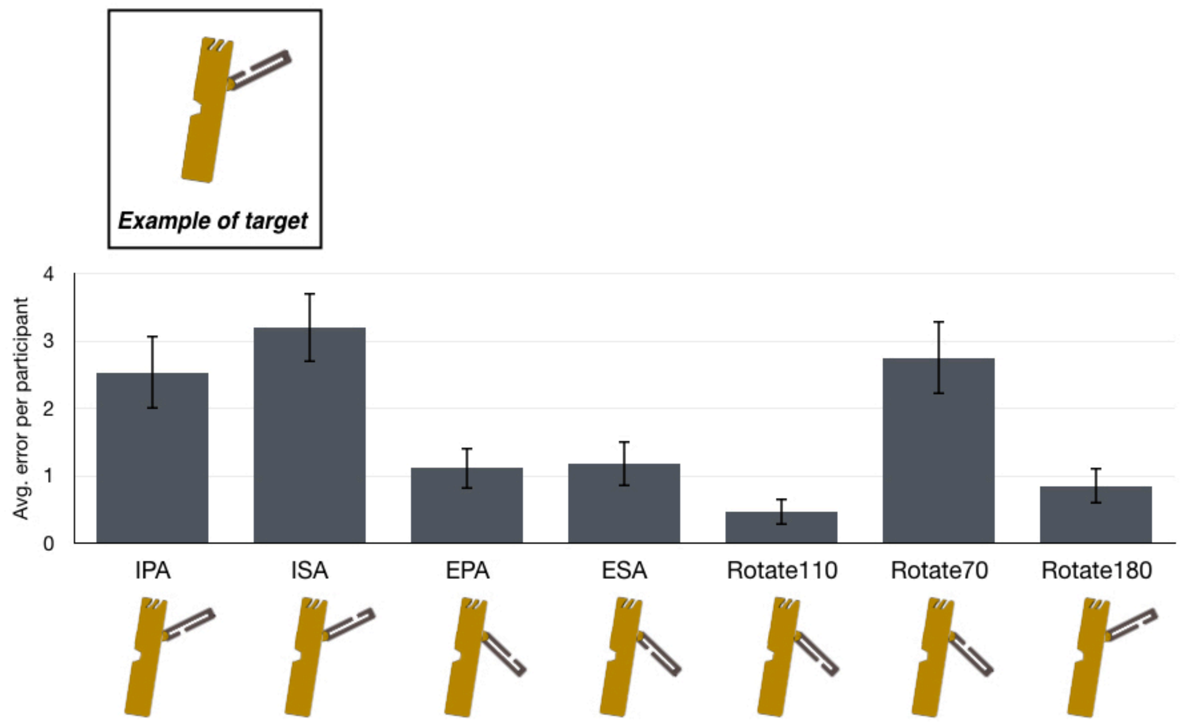

Figure 5-17. *Illustration of different types of error involving the relative orientation of the small object part.*

In addition to the IPA and the ISA errors, we also found that the errors involving a rotation of the small object part 70˚ (i.e., 70˚-rotation errors) occurred at a high rate (11.1%). Indeed, the rate of the 70˚-rotation error was comparable to the rates of the IPA and the ISA errors ($p = 1$ and 0.1, respectively), and was significantly greater than the rates of the other four types of error (i.e., EPA, ESA, 110˚-rotation errors, 180˚-rotation errors; $p \leq 0.01$, for all relevant comparisons). No other pairwise comparison was statistically significant. Further inspection of the data revealed that this pattern was robust across stimulus objects and presentation conditions (see APPENDIX A.2. Experiment 2).

Orientation Distribution for Responses with Incorrect Locations

Finally, we inspected the orientation errors of responses with incorrect locations. A response was classified as in correct if the distance from target defined at points of attachment was less than the pre-determined threshold (i.e., one-sixth the base's length). Figure 5-18A presents the orientation distribution of responses with correct and incorrect location.

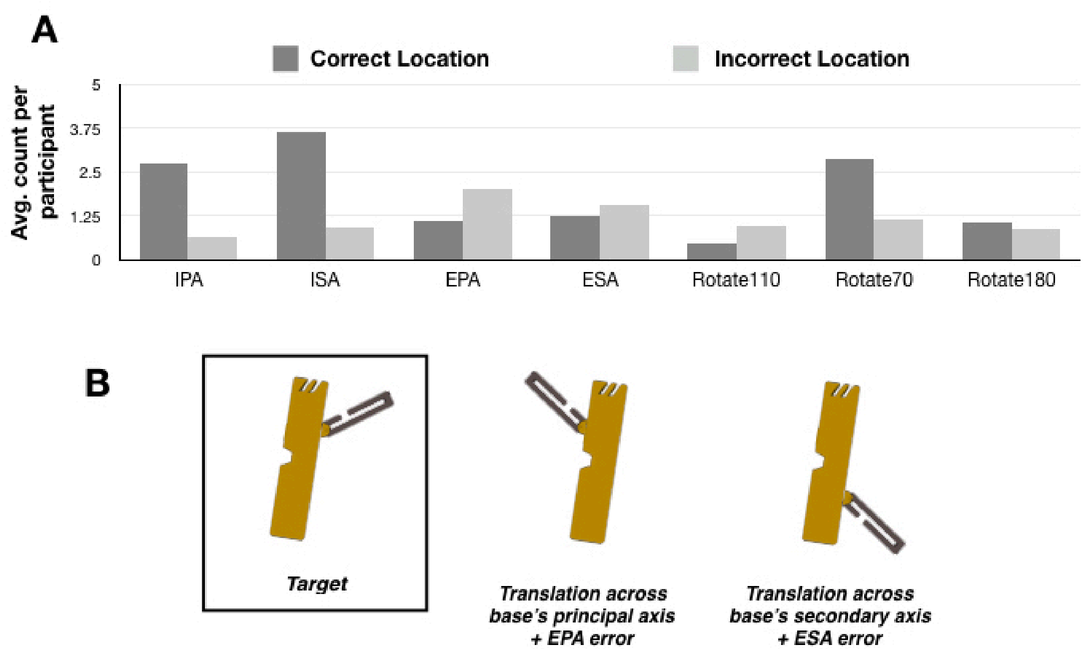

Figure 5-18. *(A) Bar plot showing orientation error distribution for responses with correct and incorrect location. (B) Illustrations for frequent types of reflection errors when the location was incorrectly remembered*

Similar to Experiment 1, the orientation distribution of incorrect-location responses (see light grey bars) was more even. However, some orientation errors were more frequent than others. A repeated-measure one-way ANOVA with Tukey-HSD post-hoc comparison revealed that the rates of the EPA and the ESA reflections were significantly greater than other error rates ($p \leq 0.05$ for all relevant comparisons). We further found that these effects were driven by some particular combinations of orientation and location error. Specifically, errors that were EPA reflections tended to correspond to a translation across the principal axis of the base (Figure 5-18B, left), and errors that were ESA tended to be a translation across the secondary axis (right). A possible explanation for these errors may be that participants occasionally misremembered the structural feature of the base. Due to lapse in attention, for instance, if they remembered the middle notch as being on the wrong side of the base, then they would produce an error corresponding to a combination of principal-axis translation and EPA error (i.e., see center of Figure 5-18B). Similarly, if they misremembered which end the jagged feature was on the base, then the error would correspond to combination of secondary-axis translation and ESA error (the right side of figure).

5.3.3. Experiment 2 Discussion

The main purpose of Experiment 2 was to examine the pattern of location errors in the shape recall task. We observed a significant occurrence of location reflection errors: When participants misremembered the location of the small object part, the responses tended to preserve the distance measured from the middle of the base part. The results can be interpreted using the framework of shape representation adapted from the COR Theory (§ 3.6.2. Representation of Relative Part Location). To represent the relative location of object parts, the brain specifies relations between sets of coordinate axes defined for the local and the global object part. This mapping requires a separate encoding of two distance parameters (i.e., *magnitude* and *direction*). If a selective disruption occurs in the encoding of the *direction* parameter, then an error involving a location reflection across an axis of the global object part is expected.

In addition to the main finding for location errors, the analyses also revealed some interesting results for when the object part's location was correctly remembered. We discuss the results and their plausible explanations below.

Basis for Defining Object Part's Location

In Experiment 2, we found that the object part's location was defined at the place of attachment: When participants correctly remembered the location but not the tilt of the object part, their responses tended to preserve the location corresponding to where the parts were attached (Figure 5-14). The current result, combined with the pattern observed in Experiment 1, suggested that the basis for defining object parts' locations could change depending on the presence of important features in the object. Specifically, adding a visible joint at the region where the parts were attached altered the basis from the center of object part to the place of attachment. The presence of the joint at the point of attachment could serve as a visual cue that makes the point easier to identify. The point that is easy to identify in a stimulus likely serves as a reliable basis for defining the object part's location, unlike a point that requires some visual estimation (e.g., the point corresponding to the center of the small part). Alternatively, the presence of the joint might alter the result because it drew participants' attention to the point of attachment.

Error Involving 70˚ Rotation

Another important finding was the unexpected high frequency of the 70˚-rotation error. When participants correctly recalled the location but not the tilt of the target stimulus in this experiment, they frequently committed an error that could be related back to the target via 70˚ rotation of the small object about the place of attachment. How might this result be explained?

Under the COR theory, the occurrence of a rotation error can be explained by assuming that participants selectively forgot the tilt direction of the orientation representation. If the knife in the figure below is remembered as being tilted 45˚ counterclockwise instead of 45˚ clockwise, for instance, then the recalled orientation will be a 90˚ rotation from the target (see also Figure 3-5 in Chapter 3).

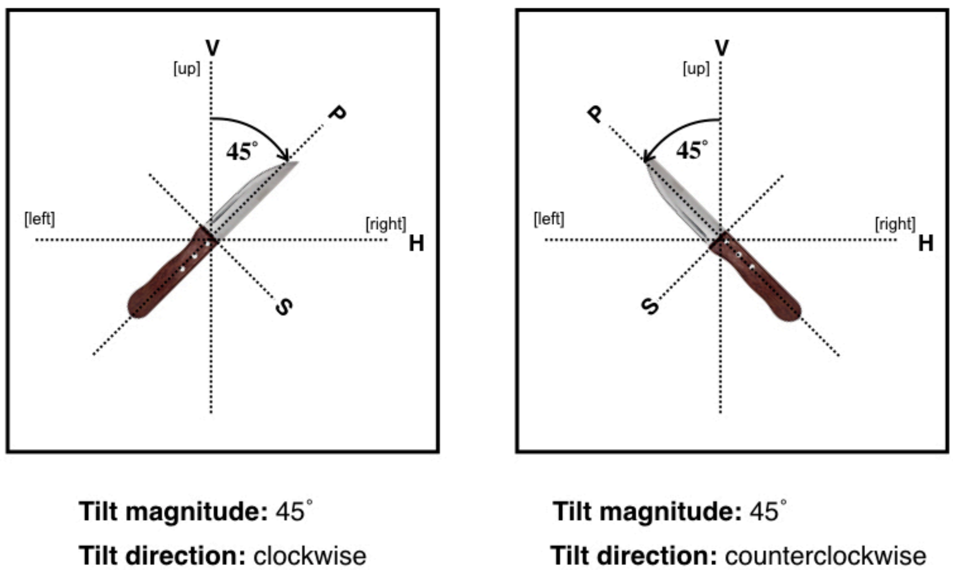

Figure 5-19. *COR representation underlying a 90˚-rotation error of a whole object.*

In representing the spatial relations of object parts, a rotation error may occur if what is forgotten is the direction of the relative tilt of object parts. As illustrated in Figure 5-20 below, the 70˚-rotation error can occur because participants forgot which direction the small part was tilted 35˚ relative to the base. We display the diagrams in this figure by aligning the origins of the object parts' coordinate axes to make the relevant angles (35˚) easy to see. This way of displaying the axes is acceptable because in orientation representation, information about the relative location of object parts (i.e., the relative location of axes' origins) is irrelevant.

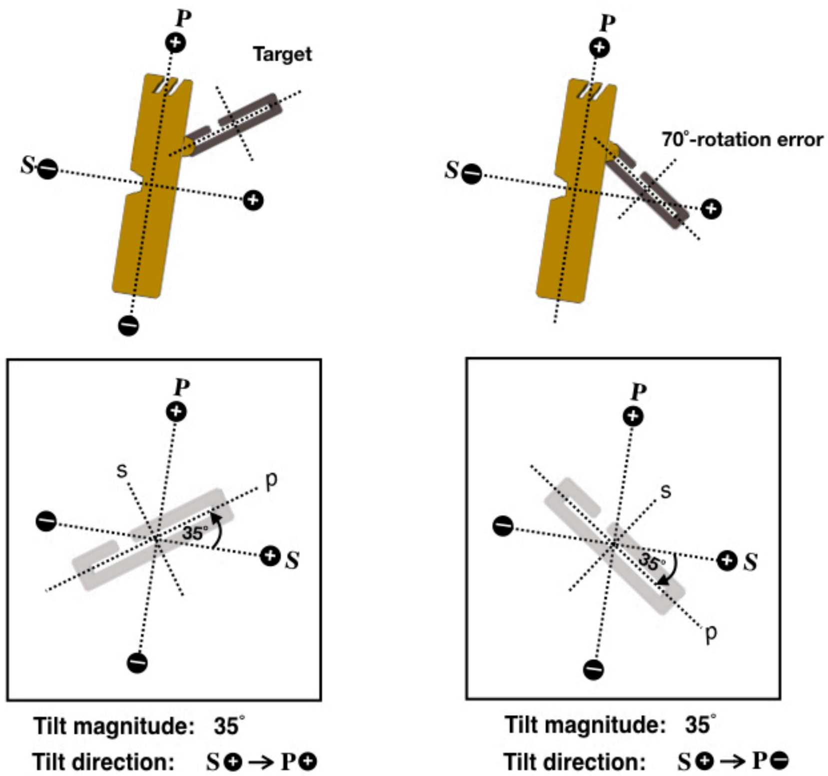

Figure 5-20. *COR representation underlying a 70˚-rotation error of an object part.*

An important question remains: Why did only the 70˚-rotation error, and not the 110˚-rotation error, frequently occur in the present experiment? To that end, why weren't the 90˚-rotation errors (i.e., 90˚FW and BW rotation) of the previous experiment also observed at high rates? One possible reason may have to do with the amount of tilt that can be encoded in an orientation representation. In the present experiment participants were always shown the small object part the angular distance of 55˚ relative to the large object part. This meant that two options of tilt representation were present: Either a representation that involve 35˚ angle (as shown in the figure above) or one that involves 55˚ angle (as shown in Figure 5-21 below). Each option required establishing spatial relations between different pairs of axes (i.e., Axis S and Axis p for 35˚ tilt, and Axis P and Axis p for 55˚ tilt). It is possible that the option that required encoding the smaller tilt was preferred. But as a result of the tilt being smaller in size, errors in recalling the tilt's direction were more common, hence the high rate of 70˚-rotation errors.

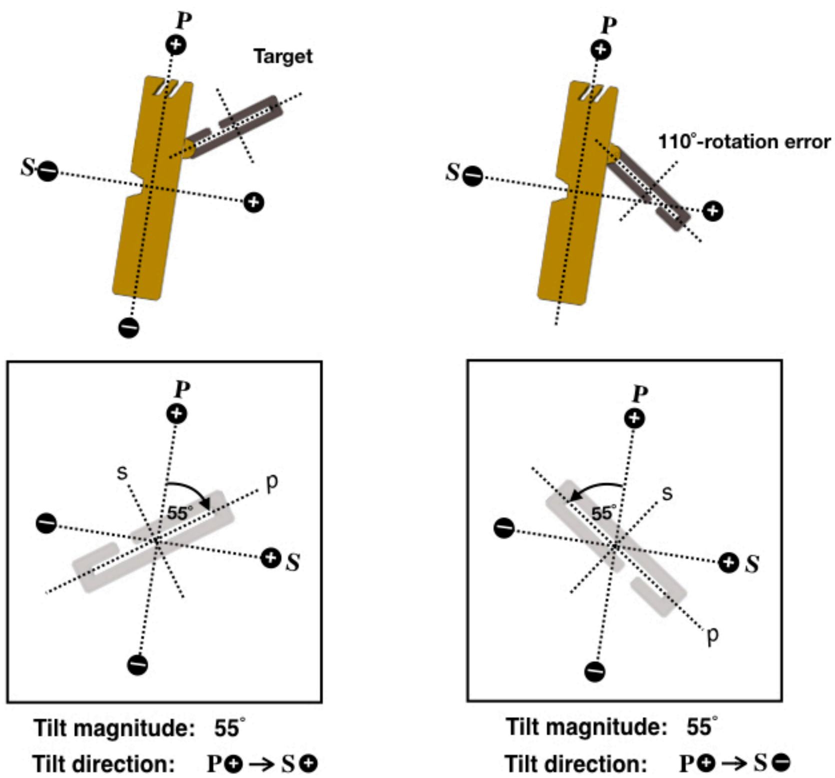

Figure 5-21. *COR representation underlying a 110˚-rotation error of an object part.*

The present explanation can also account for why we did not observe one type of rotation error occurring more frequently than the other in Experiment. In the previous experiment participants saw the small object part having 45˚ tilt relative to the base. Thus either option of tilt representation involving a different axis correspondence requires encoding 45˚. Therefore, no one option was preferred and the 90˚ rotation errors were equally common. Note that a novel COR assumption behind this explanation is that the tilt magnitude determines which pair of axes are mapped to one another. That is, the setting of the tilt parameter affects the setting of the axis correspondence parameter.

Alternative Explanations for the Orientation Pattern

Finally, we asked whether the pattern of orientation errors could be explained by the hypothesis that what the participants recalled was only the crude overall shape of the stimulus. Similar to Experiment 1, the crude-form hypothesis failed to account for why the IPA and the ISA errors occurred at higher rates compared to the 180˚-rotation error (see APPENDIX B.2. Experiment 2). In addition, the fact that the 70˚-rotation errors were observed more frequently than the EPA, the ESA and the 110˚-rotation errors, despite the fact that these errors share very similar crude features, provided yet another piece of evidence that the results of the shape recall task cannot be explained in terms of coarse-level object shape similarity. Similar to Experiment 1, we performed the CNN analysis to compare the patterns of neural network activations across different object stimuli. The results revealed a similarity pattern consistent with the prediction of the crude-form hypothesis (see bar plot in the figure).

In short, some aspects of the findings from Experiment 2 were different from the pattern observed in the previous experiment. First, when object parts' locations were correctly remembered, participants' responses often preserved the location defined at the place of attachment rather than the center of object part. Second, among these responses, errors in the form of 70˚ rotation were common. We interpreted these findings in light of the possible underlying COR mechanisms of the errors. Finally, we suggested that the new pattern of results provided yet another piece of evidence against the crude-form hypothesis.

5.3.4. Interim Conclusion

In the first two experiments, I asked whether the mechanisms for representing whole-objects' orientations posited by the COR theory can be adapted to account for the representation of object parts' orientations and locations. I hypothesized that in the shape recall task, where participants had to remember the internal arrangements of parts within an object, their common errors would be similar to the patterns observed in previous orientation studies (Chaisilprungraung et al., 2019; Gregory et al., 2011; Gregory & McCloskey, 2010). Specifically, the common errors were expected to involve orientation and location reflections of object parts. The experiments' results conformed with these predictions. In light of the adapted COR theory, I interpreted these errors as resulting from failure(s) to encode a particular component of the representation for relating the polar features (i.e., *polarity correspondence; direction*). Overall, the finding suggested that a framework for representing spatial information based on systems of perpendicular coordinate axes was useful, not only for making sense of how an entire object is oriented with respect to the environment, but also for encoding spatial relations among parts within an object. Moreover, the finding suggested that the shape recall task could be a novel valuable tool for studying object shape representation.

5.4. Experiment 3

The aim of the latter two experiments of this thesis was to examine the geometric constraints for defining axes of the overall object. As discussed in Chapter 4, the question about how the visual system defines axes in the object's reference frame was a subject of interest of many studies (Large et al., 2003; Ling & Sanocki, 1995; McMullen

& Farah, 1991; Quinlan & Humphreys, 1993; Rock, 1973; Sekuler & Swimmer, 2000). Based on available evidence, shape elongation is an important geometric property that determines how the tilts of object axes are defined (e.g., Chaisilprungraung et al., 2019; Quinlan & Humphreys, 1993; Sekuler & Swimmer, 2000). In my previous study (Chaisilprungraung et al., 2019), we found that axes of multi-part objects were defined so that the principal axis tilted in parallel with the object's elongated part (e.g., the handle of a hatchet). In a shape representation theory that assumes the notion of perpendicular coordinate axes, it is important to understand not only how the tilts of the axes are defined, but also how the *origin* of the axes (i.e., the point intersected by the object principal and secondary axes) are defined. As discussed in Chapter 4, a possible way of defining the origin of the axes is on the basis of the center of the object's elongated part (i.e., Elongated-Center Hypothesis; see Figure 4-5). Another possible way is on the basis of a point that corresponds in some way to the center of the whole object (i.e., Object-Center Hypothesis).

The latter two experiments of the thesis aimed at testing the hypotheses by using a procedure for inferring object axes similar to the previous study (Chaisilprungraung et al., 2019). In my previous study the tilts of object axes were inferred based on common orientation reflection errors (see § 3.5. Evidence for Perpendicular Coordinate Axes in Shape Representation). Participants saw a target stimulus at a random orientation on the screen, and shortly thereafter attempted to recall the object's orientation by adjusting a probe stimulus to match the target (e.g., see Figure 3-9). When a reflection error was made, we computed the axis of reflection. Across trials of the study the tilts of object axes were inferred from common tilts of axes of reflection. While this procedure was able to reveal the axes' tilts, however, it could not reveal the point intersected by the reflection axes. This was because stimulus objects were always displayed at the same location at the center of the computer screen (see Figure 4-4). Thus when orientation reflections were made, the axes of reflection always intersected at the same point, which for objects with a protruding part (e.g., a hatchet), happened to coincide with the center of the whole object Thus the previous study was unable to adjudicate hypotheses regarding the axes' origin.

The current thesis removed this limitation by freeing the locations of objects on the screen. Participants saw a target object displayed at a randomly chosen location and orientation, and then had to recall both the location and the orientation of the object by adjusting the probe to match the target (e.g., see **Figure 5-23** for an example of a trial). Reflection errors in this procedure could take place across any axis on the screen and the axes' intersections could correspond to any location on the object. In Experiment 3, I first applied this novel procedure to stimuli with simple elongated shapes (e.g., a pen, a knife). Given the simple elongated shapes of the objects, I hypothesized common points of axis intersections corresponded to the centers of the objects. The success of this experiment would provide a strong grounding for applying the procedure to more complex stimulus objects in Experiment 4.

5.4.1. Experiment 3 Stimuli and Methods

Participants

Participants were 25 Johns Hopkins University undergraduate students (16 female; 23 right-handed) with normal or corrected-to-normal vision.

Stimuli

Stimuli, shown in Figure 5-22, were photographs of 12 objects, which were also used in the first experiment of Chaisilprungraung and colleagues (2019). These photographs were taken under uniform lighting with the object axes parallel to the camera's sensor plane. All stimulus objects were chosen to be poly-oriented (i.e., regularly encountered at various different orientations), to mitigate any response bias(es) that might be associated with canonical orientations.

Figure 5-22. *Stimuli for Experiment 3.*

Procedures

The experimental procedure was similar to that of the previous study except for a few modifications. In each trial a stimulus object appeared on the screen at an orientation randomly sampled (with replacement) from the 720 possible orientations (2 enantiomorphs × 360 tilts). Participants were instructed to pay attention to how the target stimulus object was oriented. Unlike the previous experiment, the stimulus object could also appear at any of 25 locations on the screen corresponding to a 5x5 grid (see black dots in Figure 5-28 below for a complete set of possible locations[43]). Participants were instructed to remember the object's location in addition to the orientation. After the target disappeared, participants saw a probe stimulus, which was the same object displayed at a new location and orientation, and changed the location and the orientation of the probe to attempt to match them to the target's. To ensure that the task was difficult enough (and therefore that we obtained sufficient orientation reflection data), participants were tested on two target objects per trial. They first saw two target stimuli presented in succession for 750 milliseconds each (Figure 5-23A). Then, they saw the probe stimuli presented in the same order as the targets, and had unlimited time to make the response.

[43] In selecting the presentation locations, we kept in mind the size of the object, which was relatively large compared to the screen (~0.4 of screen height). To avoid presenting the object too close to an edge of the screen, we left wide margins from the screen's edges.

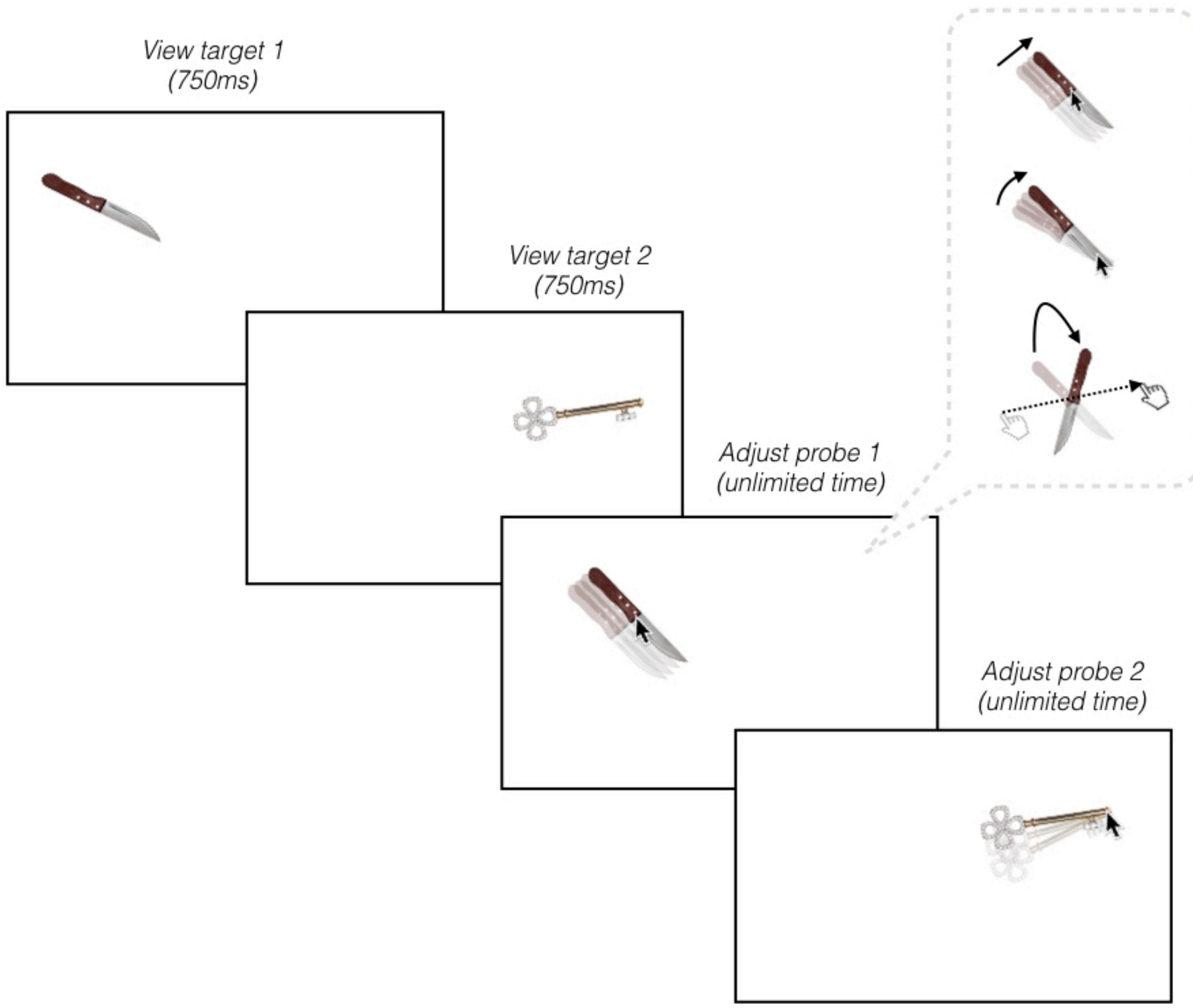

Figure 5-23. *Trial procedures of Experiment 3.*

The procedures for adjusting the probe allowed participants to change the stimulus' location and orientation using the mouse and the keyboard. These procedures were also similar to those used in the previous study, and also in Experiment 2 (see § 5.3.1. Experiment 2 Stimuli and Methods). Participants could use the mouse to drag the object to any location on the screen. Moreover, by pressing the right or the left arrow keys, they could also rotate the object in the clockwise or the counterclockwise direction. The center of the object's rotation always corresponded to wherever the cursor was placed on the screen, so that the participants had to decide in advance where to place the cursor before rotating the object. Finally, participants could flip the object by 'drawing' an axis to indicate how they wanted the object to be flipped. They right-clicked twice on the screen to mark the beginning and the end point of an axis. The object was always flipped across an axis orthogonal to the axis that was drawn. Crucially, this method of flipping let participants freely decide the axis for flipping the object, rather than forcing the flipping to take place across a pre-determined axis. The procedure thus avoided potentially suggesting that any axis had some special significance.

After the participants completed both responses, feedback in the form of accuracy score was given. Like the previous two experiments the feedback was designed to encourage participants to do well in the task without revealing how the response differed from the target. The maximum score of 100 was given if the response showed accurate location and orientation. Fifty points were allotted to the scoring of orientation accuracy, and another fifty points to the scoring of location accuracy. Responses with the maximum orientation scores were those that had the correct enantiomorph, and whose tilts were within 15˚ of the correct tilts. For each additional 30˚ departure from the correct tilt, 10 points were deduced. If a reflection error was made (i.e., a response showing the wrong enantiomorph), another 10 points were deduced. If the point deductions totaled 50 or more, the orientation score was 0. The location score was calculated based on the absolute distance from the correct answer. Participants were not informed about how the scores were calculated.

Each participant was tested on 72 target objects (i.e., six instances from each of the twelve objects), in an experiment session that lasted approximately 30 minutes. Prior to the experiment they were given extensive training. In the first six training trials, they practiced using the keyboard and the mouse to make a response. Two objects--one in semi-transparent color and one in solid color--appeared simultaneously on the screen. Participants were told to place the solid-colored object over the semi-transparent object, so that the latter was completely hidden behind the former. Then in six practice trials, they practiced recalling the object after it disappeared from the screen like in the actual experiment. Participants were constantly told that the accuracy of the response was very important. Immediately after a response was made during the practice, the correct answer in the form of semi-transparent target was presented alongside the response. If the response was incorrect, participants had to adjust the incorrect response and place it over the target until the target was completely hidden behind the response. Answer feedback was given only during practice and not the actual experiment.

5.4.2. Experiment 3 Results

Axes of Reflection

Orientation reflection errors, or the type of errors that was the main interest of the experiment, accounted for 420 of the total 1,800 responses (23.3%). For each reflection error we first computed the axis of reflection—or the axis across which the target stimulus must be reflected to yield the participant's response (see Figure 5-24A). Then, we plotted the reflection axis as if the target had been presented at the center of the screen (Figure 5-24B), and at a predesignated vertical orientation, which we refer to as the normalized orientation (Figure 5-22 shows the stimulus objects at their normalized orientations). In this coding scheme the center of the elongated object corresponds to the center of the computer screen. Thus, if object axes intersect at the center of the elongated object, the point of intersection should correspond to the point at (0,0) on the screen.

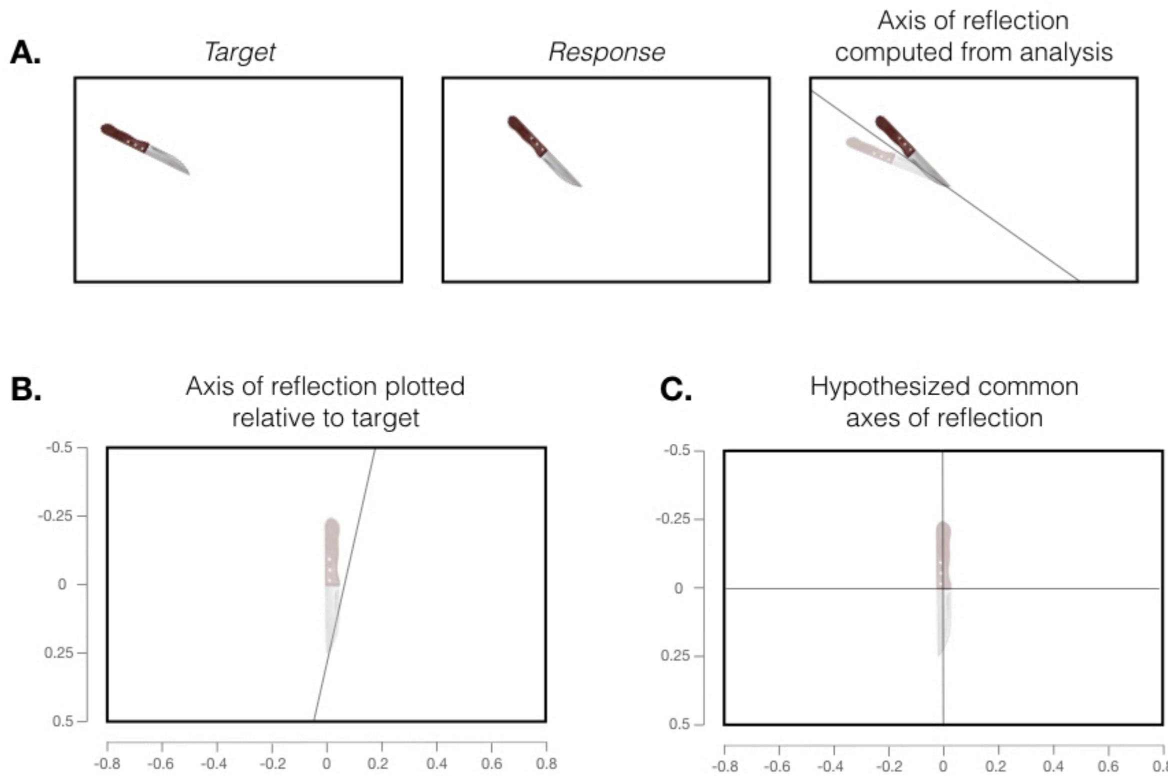

Figure 5-24. *(A) Example showing how an axis of reflection was computed based on a pair of a target and a response, and (B) how it was plotted. (C) Predicted common axes of reflection.*

The reflection-axis data are displayed in Figure 5-25A. In this figure we display the reflection axes by superimposing them over the objects' bounding boxes--that is, the smallest boxes that could contain the objects (see green boxes in the figure). As the plot suggests, most reflection errors took place across an axis that was parallel or nearly parallel with the object's axis of elongation. Also somewhat common were errors with reflection axes approximately 90˚ of the object's axis of elongation. In a separate analysis, we exclusively examined the orientations of the reflection axes to see whether the axes' tilts replicated the pattern observed in the previous study (Chaisilprungraung et al., 2019). Like the previous study, we computed the 'best-fit' axes--or the pair of perpendicular axes that minimized the sum of absolute angular distances to the individual reflection axes (see Figure 4-4C in Chapter 4). The result (see APPENDIX C.1. Experiment 3) revealed that the best-fit axes had the orientations of -1˚ and +89˚, suggesting that the tilt pattern of the reflection axes was the same as that observed in the previous study. The result also provided a grounding for inferring the principal and secondary object axes based on the reflection axes.

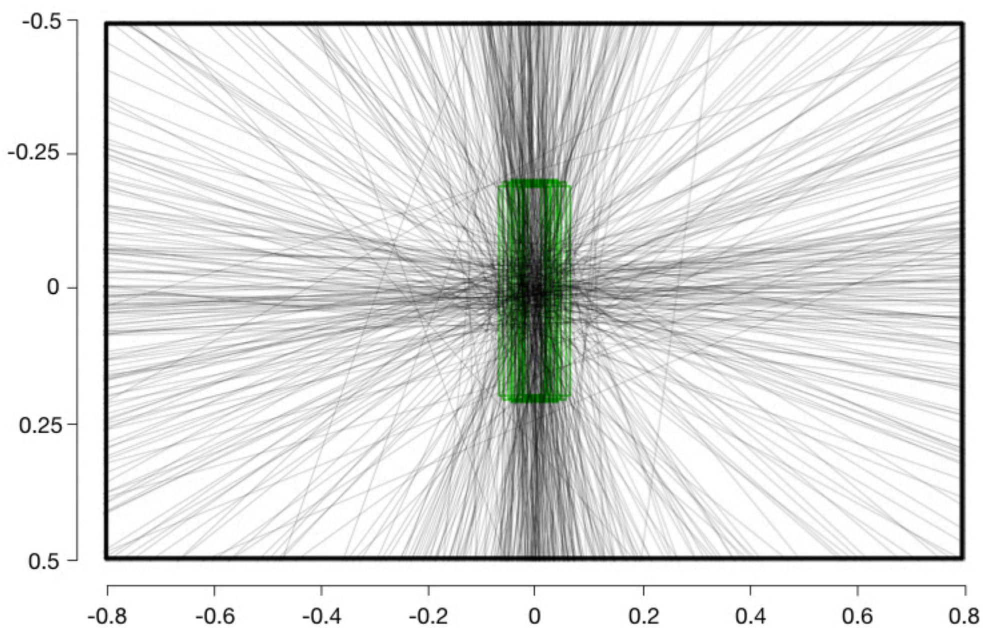

Figure 5-25. *All axes of reflection in Experiment 3.*

The plot in Figure 5-25 suggests that most axes' intersections occur at a location very close to the center of the elongated object. The analysis of axis intersections was performed separately for each stimulus object (Figure 5-26A). For each object, we selected axes of reflection that we assumed corresponding to either the principal or the secondary object axis--that is, axes having tilts within 15° or $90° \pm 15°$ (in clockwise or counterclockwise direction) of the object's elongation. We identified all points of intersection formed by the axes of reflection, and plotted the mean location of the points of intersection (see black dot in the figure) for each object.

Figure 5-26B summarizes all mean locations of the intersection in one plot. Across 12 objects, the average location of intersection was -0.0021 on the X-axis, and -0.0019 on the Y- axis.[44] In other words, the X- and Y- distances of the intersection from the center of the elongated object (i.e., defined as the point at (0,0)) was approximately 0.2% of the height of the computer screen. Single-sample t-tests revealed that the distances were not significantly greater than zero (i.e., $t(11) = 0.80, p = 0.4$ for distances on the X-axis and $t(11) = 0.18, p = 0.9$ for distances on the Y-axis). The analysis therefore suggested that the point of intersection tended to correspond to the center of the elongated object, consistent with the experiment's prediction. Further analyses revealed that this pattern was consistent across locations where target stimuli were presented on the screen (see APPENDIX A.3. Experiment 3).

[44] The unit of measurement is the height of the computer screen, so that the value of 1 corresponds to a location away from the center of the screen in the distance of the screen's height. We also tried scaling this location relative to the object itself. Relative to the object's width, the average location of intersection was -7.9%, and relative to the object's height, the location was -0.4%.

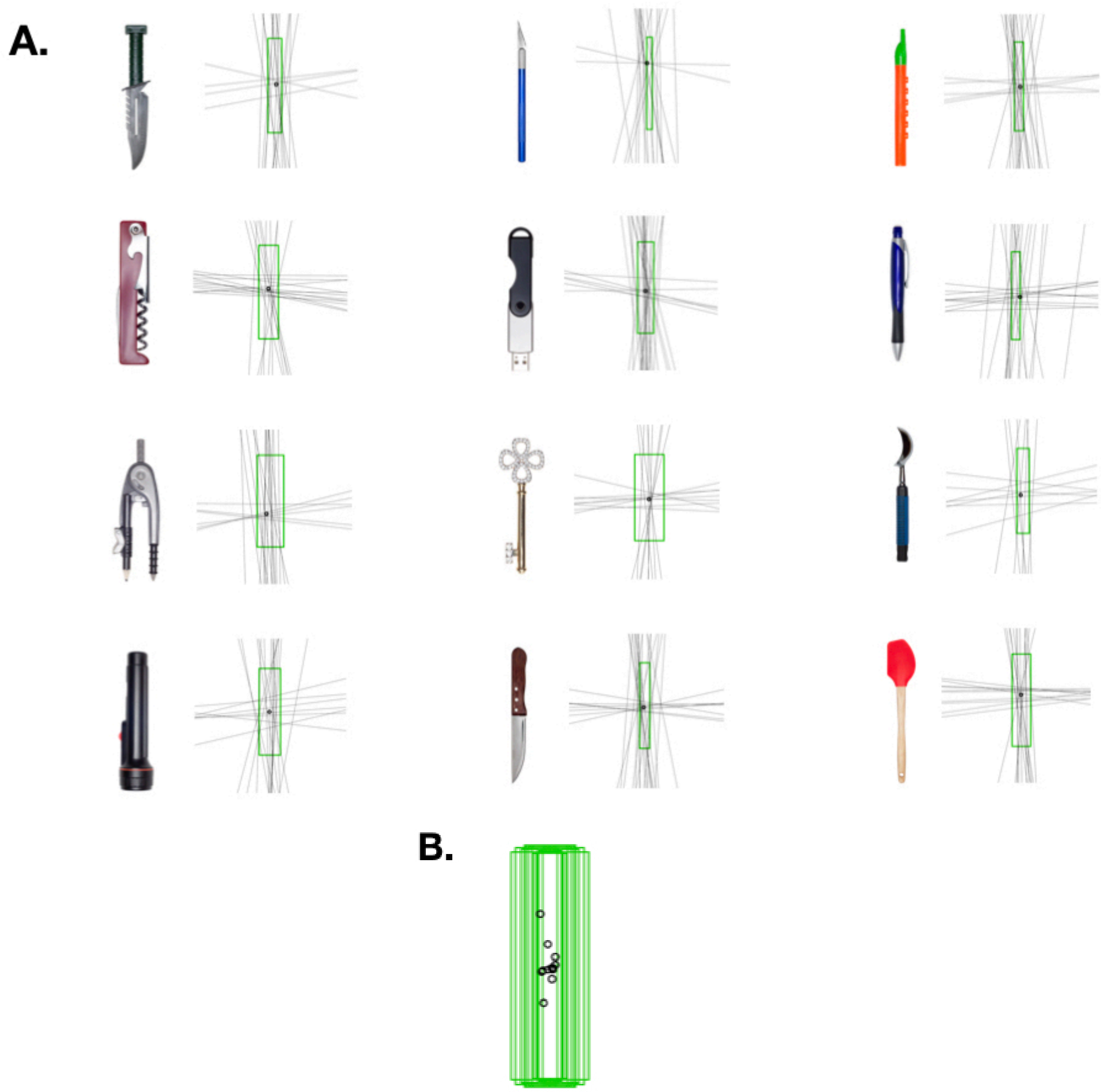

Figure 5-26. *(A). Axes of reflection for each object of Experiment 3. (B) Mean locations of the axes' points of intersection.*

Tilt and Location Accuracy

In addition to the main analysis on reflection data, the experiment also examined participants' responses when the object was not reflected (i.e., when the enantiomorph of the target stimulus was correctly remembered). These non-reflection responses accounted for 1,380 of the total 1,800 responses and could occur in the form of an incorrect tilt in the picture plane, and/or an incorrect recall of the object's location. Below, we separately examined the tilt and the location accuracy of the objects.

For the tilt accuracy, we plotted the distribution of the angular differences between target and response orientations (Figure 5-27). The median absolute error of the differences was 9.8˚, indicating that participants were reasonably accurate at reproducing the tilt of the target stimuli. Also, the distribution showed no overall bias toward clockwise or counterclockwise tilts from the correct orientation: The circular median of the signed angular distances was 0.04˚.

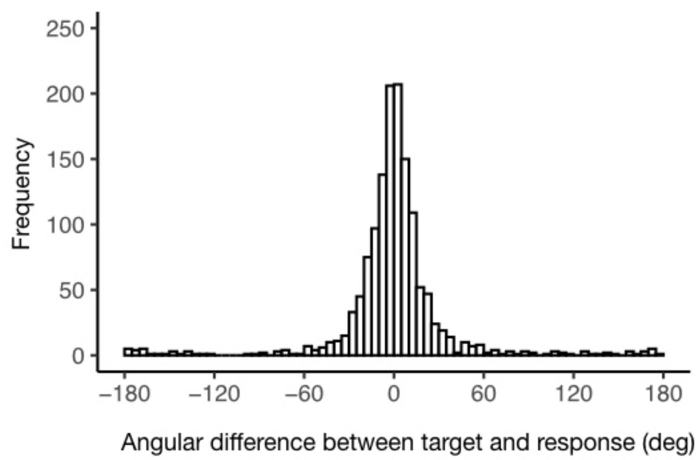

Figure 5-27. *Histogram for non-orientation reflection responses in Experiment 3.*

For the location accuracy, two types of analysis were possible based on how response locations were defined. In the first analysis, the locations were defined relative to the target at the pre-designated orientation and location (see semi-transparent circles Figure 5-28A). This was similar to how reflection axes were defined in the previous analysis. The mean location of the response was -0.002 on the X-axis (SD=0.085), and

0.012 on the Y-axis (SD=0.06). suggesting that participants were reasonably accurate at remembering the object's location on the screen.

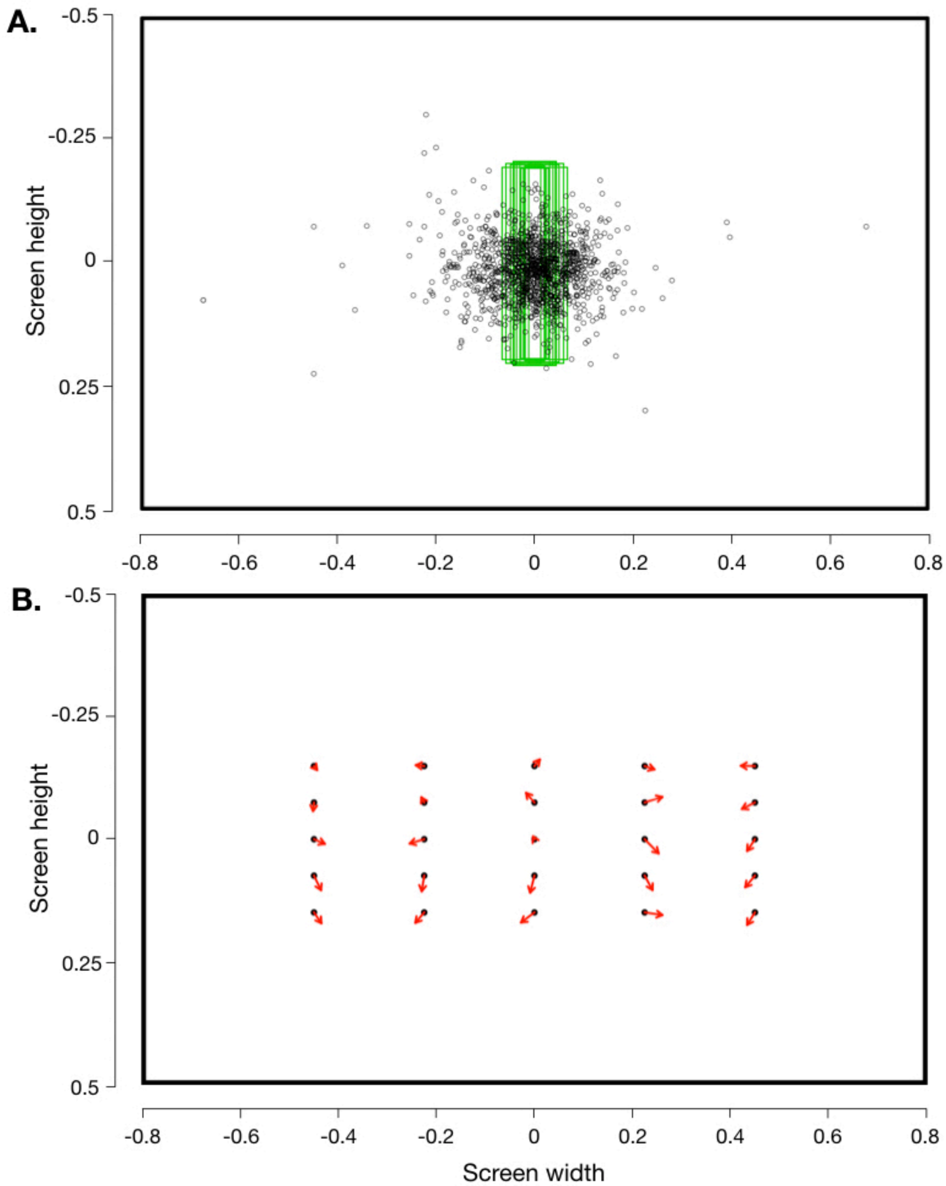

Figure 5-28. *Plot showing the location accuracy of participants' responses in Experiment 3. (A) Responses plotted relative to the target's locations on the screen. (B) Responses plotted as if the target was presented at a vertical orientation at the middle of the screen.*

Another way of defining the response locations was relative to the target as it was actually presented on the screen (i.e., relative to the target at is actual orientation and location of presentation). In Figure 5-28B, each black dot denotes one of 25 possible locations of target presentation on the screen. The average displacement of participants' responses from each location of presentation is depicted with a red arrow. The length of the arrow corresponds to the average amount of displacement, and the direction of the arrow the average direction of the displacement. Across all 25 locations of presentations, the average amount of displacement was 0.012 (or approximately 1.2% of the screen's height[45]), and the average direction was towards the bottom of the screen and slightly to the left (i.e., 9.6˚ clockwise if the downward direction is defined as 0˚). Together, the results suggested that participants made minor translation errors in recalling the location of the target stimuli, and that no bias in direction was present, either when the translations were defined relative to the object or relative to the screen.

5.4.3. Experiment 3 Discussion

Experiment 3 asked whether the procedure for inferring object axes from reflection patterns could be modified to also probe the origin of object axes. In a task where participants had to recall both the orientation and the location of a stimulus object on screen, we found that common points of axis intersections corresponded to the center of the elongated object. The analyses on the non-reflection responses suggested that participants were generally accurate at remembering the tilts and locations of target stimuli. The results conformed with the predictions of the experiment, and potentially suggested that the new procedure could be used to infer the origin of object axes.

---

[45] For comparison, the length of a stimulus object was 40% relative to the screen's height.

Before applying the procedure to Experiment 4, however, some implicit assumptions of the procedure deserve further discussion. In inferring object axes based on reflection patterns, we assumed that the reflection errors occurred primarily because participants forgot how the polar features of the axes were related. In the COR theory, if a representational failure affects the coding of the axes' polarity mapping, but leaves all other representations intact (e.g., tilt), then a reflection error across an object axis is expected (see § 3.4.1. Coordinate Orientation Representation (COR) Theory). Reflection errors that occurred purely as a result of an incorrect polarity mapping take place across an axis that perfectly aligns with either the object principal axis or the object secondary axis. Therefore, object axes may be directly inferred from the axes of reflection.

Figure 5-29. *Illustrations showing an example of how a combination of different representational failures can give rise to a reflection response.*

In practice, however, participants' errors rarely occurred because they *only* forgot how the axes' polarities were related. Due to noise in the encoding or the storing of visual information, participants likely also misremembered exactly where the object was presented on the screen, or how it was tilted. Figure 5-29 illustrates a hypothetical example where a reflection response occurred due to a combination of a polarity mapping error, and errors in recalling the exact tilt and location of the object (see the right panel of the figure). The red double arrow depicts the axis of reflection, which we defined in the experiment as the axis across which the target stimulus must be reflected to yield the response. As the example illustrates, the axis of reflection provides a loose basis for inferring the object axis. This is because the response occurs as a combination of multiple representational failures, rather than a pure reflection error.

The current experiment addressed this problem by making certain assumptions regarding the distributions of the tilt and the location errors. Particularly, we assumed that the tilt and location inaccuracies were present in individual trials in small amount, and that across trials, there was no systematic bias in the direction of the errors (e.g., whether the object was remembered as being closer to one side of the screen than it actually was). Because the mean directions of the location and the tilt errors were assumed to be close to zero, it was reasonable to expect the average reflection data to provide a good estimate of the 'true' reflection axes—or axes of the reflection errors where no tilt or location error was made (i.e., when the minor errors were canceled out). In our analysis of non-reflection responses, we found that participants were generally quite accurate at remembering the tilt and the location of the stimuli on the screen, and that across trials, there was no systematic bias in the direction of the errors (Figure 5-26 & Figure 5-27). Therefore current evidence supported our assumption regarding the error distribution, suggesting that the origin of object axes may be inferred based on the average point of intersection.

However, it is not always possible to expect that the assumptions regarding the error distribution will be met. More generally, it is helpful to understand whether and in what way the reflection-axis pattern is affected if a large bias in the tilt or the location error was present. In a set of additional analyses (see Appendix D for full details), I created simulated responses of participants by manipulating some underlying assumptions about the distribution of the tilt and location errors. In one simulation, reflection data was constructed based on a hypothetical assumption that participants were grossly inaccurate at remembering the location of a target stimulus on the screen (Figure D- iiA). In the other two simulations, the analysis assumed that participants systematically remembered the target stimuli as being tilted towards a particular direction (Figure D- iiB) or being presented nearer to a particular side of the screen than they had been shown (Figure D- iiC). In all simulations, all other aspects of the simulated responses (e.g., rate of reflection error, shape of error distribution) were kept as close to the experimental data as much as possible. I found that the axis intersections invariably took place at the center of the elongated object regardless of the manipulations of the tilt and location errors (see right panel of Figure D- ii). The simulation therefore suggested that the procedure could be used to infer the origin of object axes even when large biases in tilt or location were present.

5.5. Experiment 4

Experiment 4 applied the procedure for inferring the origin of object axes to stimuli that consisted of an elongated part and a protruding part. In this experiment, we plotted the axes of reflection as if the target object had been displayed with its elongated part at the center of the screen (see Figure 5-30). If the axes' origin is defined based on the elongated part of the object (i.e., Elongated-Center Hypothesis), the point of intersection of the reflection axes should frequently correspond to the point at (0,0) (see left side of the figure). If the axes' origin is defined based on the object's overall center defined in some way (e.g., the center of the smallest box that can contain the object; Object-Center Hypothesis), the average point of axis intersection should shift in the plot towards the direction of the object's protruding part (see right side of the figure).

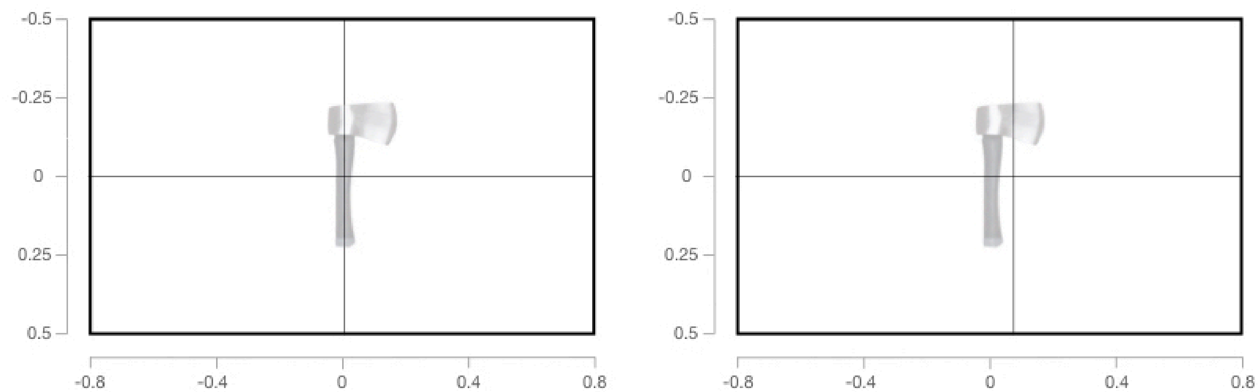

Figure 5-30. *Predictions for hypotheses of Experiment 4: The Elongated-Center Hypothesis (left) and the Object-Center Hypothesis (right).*

5.5.1. Experiment 4 Stimuli and Methods

Unless explicitly noted, the method of Experiment 4 was the same as that of Experiment 3.

Participants

Participants were 45 Johns Hopkins University undergraduate students (30 female; 41 right-handed) with normal or corrected-to-normal vision.

Stimuli and Procedure

Stimuli were photographs of 12 objects with a protruding part attached to the end of an elongated part at an angle of approximately 90˚ (Figure 5-31). These stimulus objects were also the same as those used in the second experiment of Chaisilprungraung and colleagues (2019).

Figure 5-31. *Stimulus objects for Experiment 4.*

Testing procedures were similar to Experiment 3, except that participants practiced recalling the target stimuli on seven instead of six trials. The addition of one practice trial was in order to give participants more familiarity with the task prior to beginning the experiment. To account for the increased time spent during practice, and also for the fact that participants generally spent longer time making a response in this experiment, we reduced the number of experimental trials from 36 to 30 (i.e., 60 target stimuli per participant).

5.5.2. Experiment 4 Results

Axes of Reflection

Orientation reflection errors accounted for 495 (i.e., 18.3%) of the total 2,700 responses collected. A plot containing all axes of reflection is provided in Figure 5-32 below. Like Experiment 3, we first examined whether the tilts of the reflection axes conformed to the pattern observed in the previous study (Chaisilprungraung et al., 2019). The results (see Appendix C.2. Experiment 4) revealed the best-fit orientations of 5˚ and -85˚.

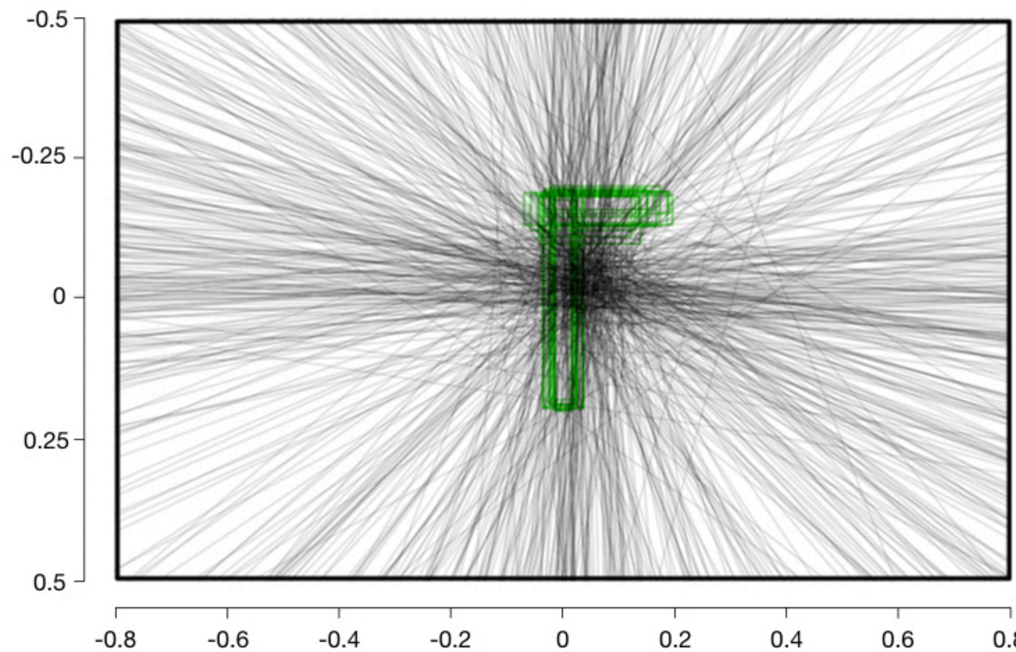

Figure 5-32. *All axes of reflection in Experiment 4.*

In Figure 5-32, the axes of reflection are displayed over green boxes which had been manually drawn to correspond to the dimensions of the object parts (Appendix E shows images of manually drawn boxes directly superimposed over the objects). Like the previous experiment, we computed the axes of reflection and the mean location of intersection separately for each object (Figure 5-33A). Since it was assumed that object axes were representationally defined on the basis of the object's elongated part, the reflection axes included in the analysis were those having tilts within 15˚ or $90^\circ \pm 15^\circ$ of the object's axis of elongated part. Figure 5-33B shows the average points of axis intersection for the different objects. Across twelve objects, the overall mean location of intersection (see red circle in the figure) was 0.03 along the X-axis, and -0.04 along the Y-axis.[46] Single-sample t-tests revealed that the distance between this mean location and the center of the elongated part (i.e., the point at (0,0)) was significantly greater than zero (i.e., $t(11) = 4.42, p = 0.001$ for X-axis distance and $t(11) = 8.06, p = 0$ for Y-axis distance).

[46] Relative to the object's width, the average point of intersection was 20.4% and relative to the object's height, the point was -8.1%

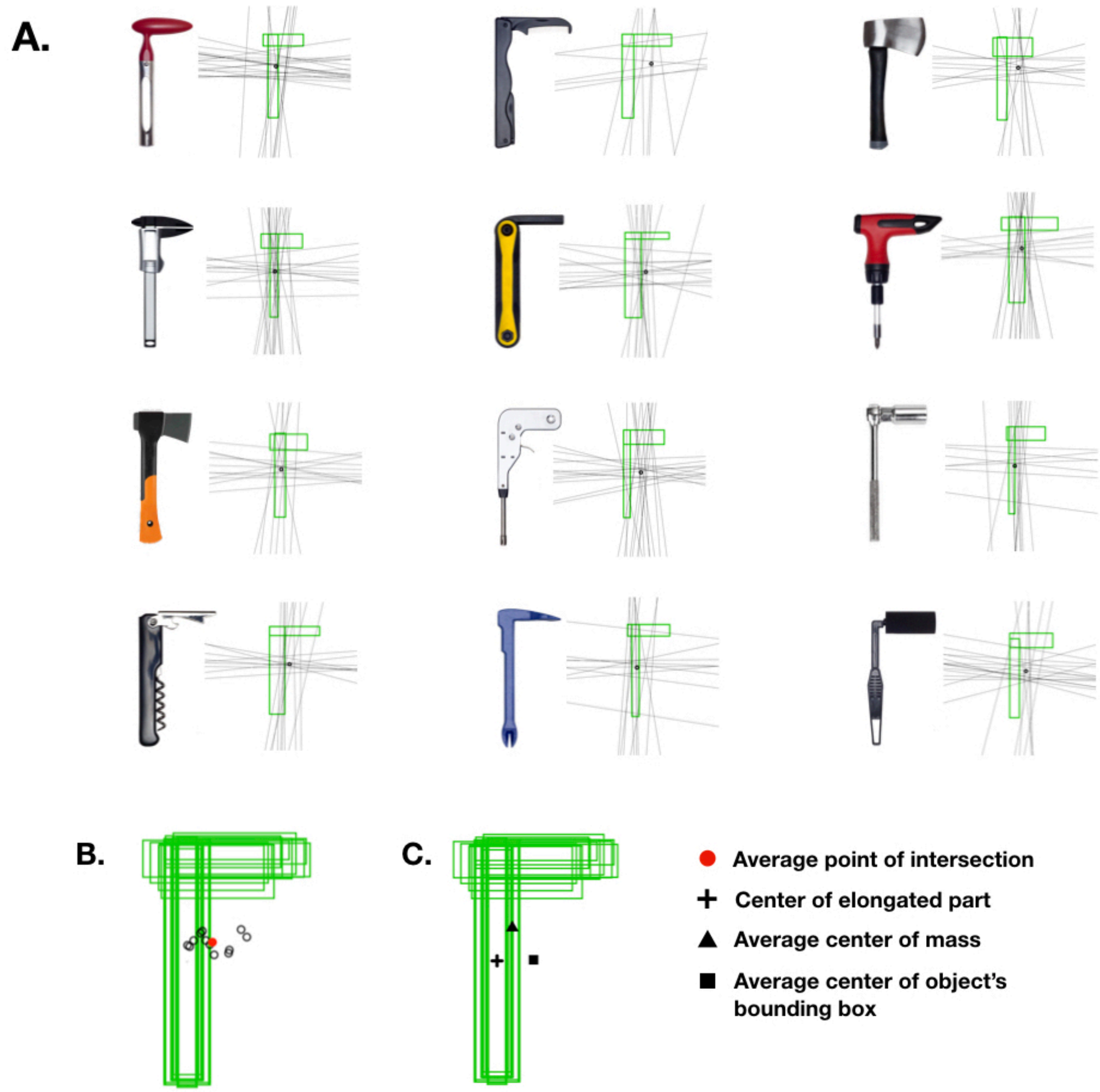

Figure 5-33. *(A). Axes of reflection for each object of Experiment 4. (B) Mean locations of the axes' points of intersection.*

The result obtained from the analysis above is interesting. While the mean location of axis intersection significantly differed from the center of the elongaed part, it neither fully conresponded to the overall object's center as defined by the center of the smallest box that could contain the object (i.e., the bounding box). We computed the average center of the bounding boxes (see black square in Figure 5-33C). Single-sample t-tests revealed that the distance between the axes' intersections and the average center of the bounding boxes was significant along the Y-axis ($t(11) = 7.6, p = 0$), and marginally significant along the X-axis ($t(11) = -2.1, p = 0.06$). In addition to the centers of the bounding boxes, we also computed the center of mass (i.e., the centroid) for each object. The average centroid for all 12 objects is depicted as a black triangle in Figure 5-33C. Single-sample t-tests revealed that this average centroid significantly differed from the points of intersection along the Y-axis ($t(11) = -5.9, p = 0.0001$), but not the X-axis ($t(11) = 1.7, p = 0.1$). Further analyses revealed that similar results were observed across locations of target presentation on the screen (see Appendix A.4. Experiment 4). Together, the results suggested that the axes' origin was not defined on the basis of the object's elongated part, inconsistent with the Elongated-Center Hypothesis's prediction. The results also indicated that axes' interesections occurred near a location corresponding to the center of the overall object. However, available evidence is inconslusive regarding how this center of the overall object may be defined.

Tilt and Location Accuracy

Like Experiment 3, we examined participants' responses when the enantiomorph of the target stimulus was correctly remembered. These non-reflection responses accounted for 2,205 of the total 2,700 responses. Figure 5-34 below presents the distribution of the angular differences between target and response orientations. The median absolute error of the differences was 12.7˚, suggesting that participants were reasonably accurate at reproducing the tilt of the target stimuli. Also, similar to the previous experiment, the distribution showed no overall bias toward clockwise or counterclockwise tilts from the correct orientation: The circular median of the signed angular distances was 1.2˚.

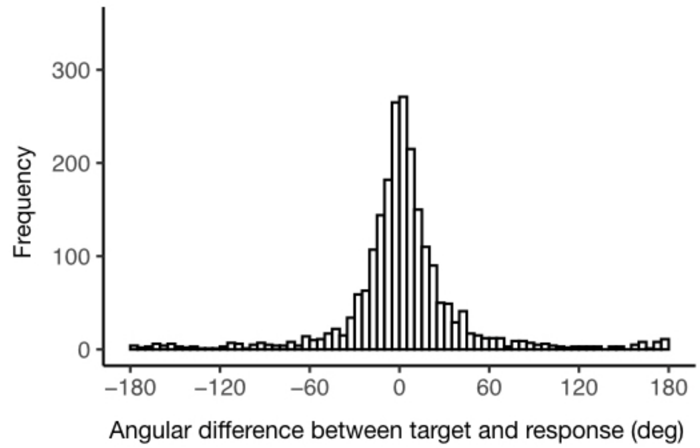

Figure 5-34. *Histogram for non-orientation reflection responses in Experiment 4.*

We also examined accuracy of the locations of participants' responses. Figure 5-35A shows responses (i.e., black circles in the figure) plotted relative to the target at the pre-designated orientation and location. The mean location of the response was -0.001 on the X-axis (SD=0.09), and 0.009 on the Y-axis (SD=0.06). suggesting that participants were reasonably accurate at remembering the object's location on the screen.

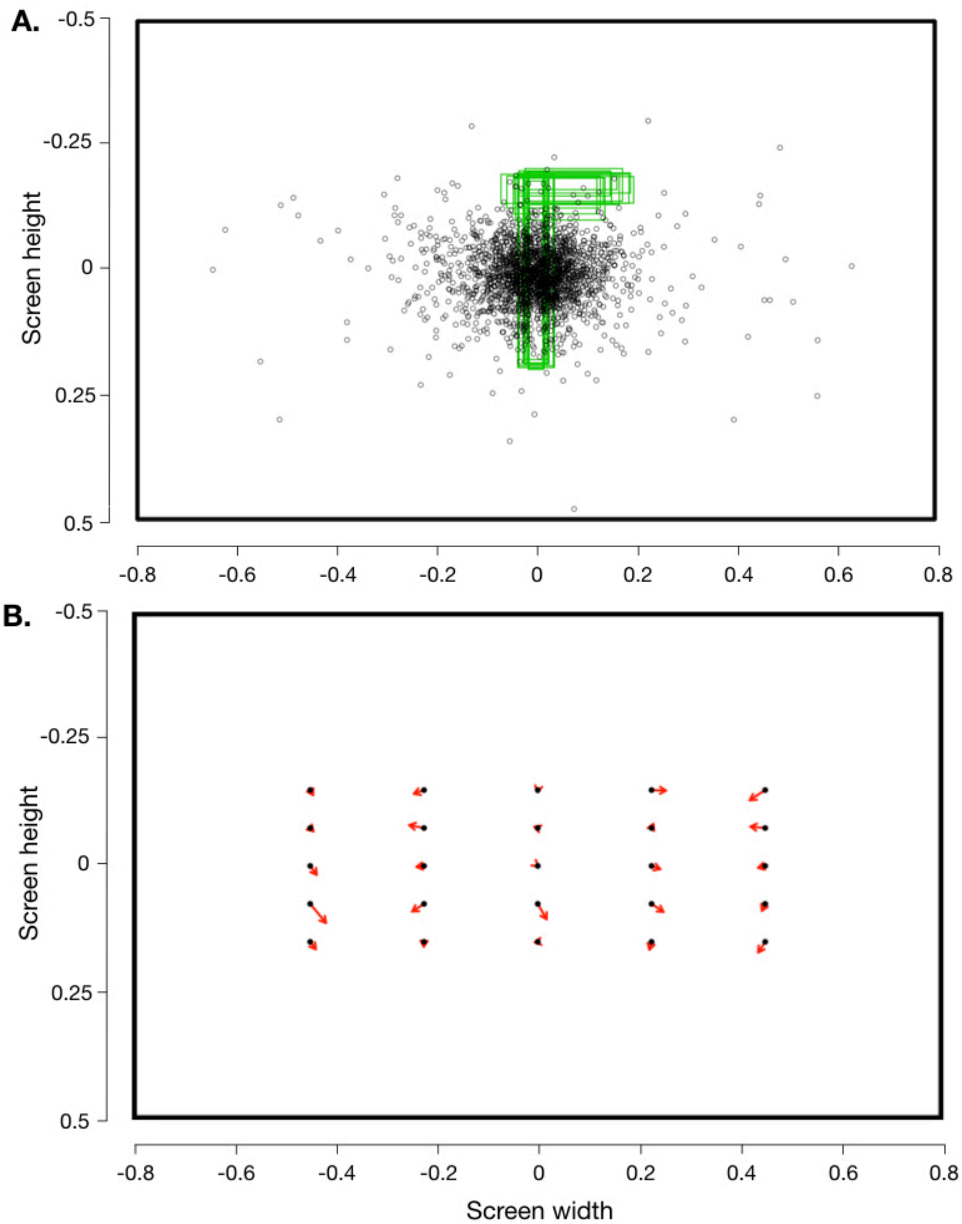

Figure 5-35. *Plot showing the location accuracy of participants' responses in Experiment 4. (A) Responses plotted relative to the target's locations on the screen. (B) Responses plotted as if the target was presented at a vertical orientation at the middle of the screen.*

Figure 5-35B shows participants' responses defined relative to the target as it was actually presented on the screen. The average distance and direction of a response's displacement is shown in the form of an arrow. Across 25 locations of presentations, the average amount of displacement was 0.009 (or approximately 0.9% of the screen's height), and the average direction was towards the bottom of the screen and slightly to the left (i.e., 6.4˚ clockwise if the downward direction is defined as 0˚). Together, the finding suggested a similar accuracy in recalling the object's tilt and location as in the previous experiment. In addition, no large bias was found in the direction of the tilt and the location error.

5.5.3. Experiment 4 Discussion

In Experiment 4 we found that the average point of axis intersection did not correspond exactly to the center of the object's elongated part as would be expected under the Elongated-Center Hypothesis. Rather, the average intersection took place closer to the center of the overall object. For convenience, the current experiment chose the center of mass and the center of the object's bounding box (Figure 5-33) to illustrate examples of how the center of the overall object may be computed.

It is important to stress that the center of mass and the center of the bounding box were only chosen as examples. The fact that the current result was better explained in terms of these locations does not necessarily imply that the way the visual system computes the axes' origins is precisely the same as how the locations are mathematically defined. For example, the center of mass is defined as the arithmetic mean location of all points (i.e., mass) in a shape. For a two-dimensional image, defining the center of mass requires treating the image as binary pixels. The current finding does not necessarily suggest that this is the way the visual system computes origins of object axes. For example, the center of the overall object may be computed by assigning pre-determined weights to different parts of the object (e.g., 0.2 for the small object part and 0.8 for the large object part), and calculating the average position of the object parts' centers based on the assigned weights (Cohen & Singh, 2006). Regardless of precisely how axes' origins are computed, nonetheless, the current experiment suggests that the computation must involve taking all parts or all regions of an object into account.

What might the current finding reveal about the critical properties of shape geometry that determined object axes? In addition, how might the current result be reconciled with our previous finding that the tilts of object axes were determined based on the object's elongated part (Chaisilprungraung et al., 2019)? A possible answer to these questions may require considering possible outcomes of changes resulting from movements of object parts relative to one another. In the previous study we suggested that axes' tilts were determined based on an individual object part (as opposed to the overall object) because such axes were invariant to minor shift in position of parts within the object. If the axes of the human body are defined based on the direction of the torso (e.g., see Marr's illustration; Figure 3-3A), then the tilts of the axes need not be redefined or recomputed every time there is a shift in position of the limbs in the body. Critically, the invariance of the axes' tilts also means that the axes' polar features (e.g., the top vs. down and left vs. right features of the human body) remain the same regardless of changes in the body's parts.

The constancy of the axes' polar features, however, does not require that the point intersected by the object axes (i.e., the origin of the coordinate system) also be invariant over movements of parts relative to one another. As an illustration of this point, consider the axes defined for the bottle opener in Figure 5-36. In this figure the tilts of the object axes are defined based on the object's elongated part (i.e., the black plastic piece that serves as a handle). The origin of the object axes, however, is defined based on the object's global shape.[47] As a consequence of this way of defining the axes, only the axes' origin changes with the relative movement of object part (i.e., whether the metal part is unfolded or half-folded). Because the tilts of the axes are invariant over the movement, the axes' polar features (i.e., [straight] vs. [protruding]; [metal] vs. [butt]) remain constant.

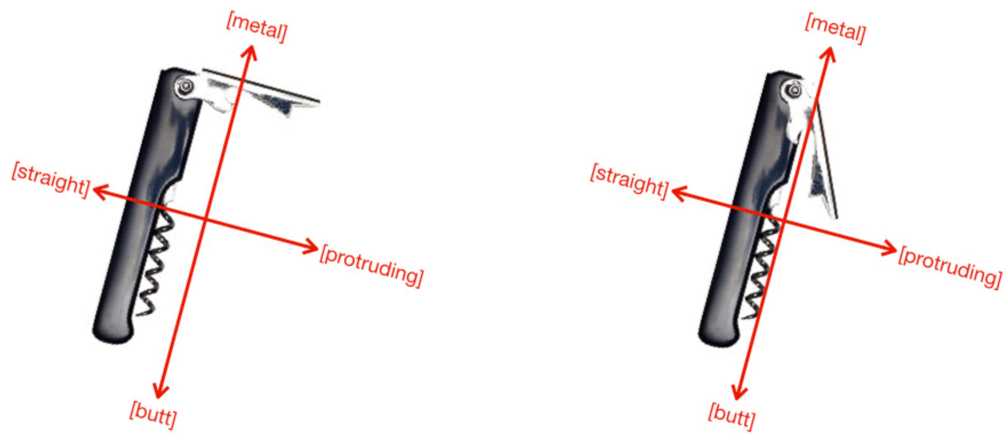

Figure 5-36. *Illustrations of axes for a bottle opener when the object is unfolded (left), and when the object is half-folded (right).*

The constancy of axes' polar features may be important in shape processing in several ways. According to some theoretical perspectives, axes' polar features are necessary in object recognition. To recognize an object, the brain creates a perceptual representation for the object's shape, and then matches it to the canonical shape representation stored in long-term memory (Hock & Tromley, 1978; Jolicoeur, 1985;

[47] For purpose of illustration, we use the center of the object's bounding box to display the axes' origin.

Jolicoeur & Humphrey, 1998; Lawson, 1999; Leek, 1998; McMullen & Farah, 1991; Rock, 1973, 1983; Tarr & Pinker, 1990). However, because real-world objects are not always encountered at their usual orientations (e.g., a tilted teapot, a fallen tree), a process akin to mental rotation has to take place. In this view the transformation proceeds until a match is found for the axes' polar features across the perceptual representation and the canonical representation.[48] If the polar features are determined in the same way regardless of how parts in the object are positioned relative to one another (e.g., bottle opener with the metal piece folded vs. unfolded), then recognition can be carried out regardless of the object parts' movement.[49]

Another possible way to explain the current result is to consider possible roles of the origins of object axes. In the same way that the origin of an object part serves as a basis for representing the part's location (see § 3.6.2. Representation of Relative Part Location), the origin of the overall object may also provide a basis for representing where the entire object is in the environment. To represent where a chair is in the room, for example, the visual system may encode the chair's location defined at the origin of the chair's axes with respect to the reference frame of the room. The origin that is computed by taking all parts of the chair into account (i.e., legs, seat, back post) may provide a

---

[48] Support for this view came from the finding that the time it took participants to recognize an object at an unfamiliar orientation in the picture plane (e.g., a tree rotated 120˚ from upright) often increased as a function of the orientation disparity from the object's usual or most familiar orientation (Jolicoeur, 1985; Lawson, 1999; Leek, 1998; Maki, 1986; McMullen & Farah, 1991; Rock, 1973; Tarr & Pinker, 1990).

[49] Indeed, the same can be said for variations that occur due to changes in length or size of individual object part. If the metal piece of the bottle opener were to be slightly shorter or longer, for instance, the shape can still be recognized as the opener so long as the polar features are constant.

better point of reference for the object's location, because it better captures the overall object shape.

Finally, it may be asked why the elongated part serves as a good basis for defining the axes' tilt--and thereby, a basis for characterizing intrinsic directions of the object's reference frame. One possible reason may have to do with the fact that in most real-world objects, the elongated part is often also the most structurally central part--the part that, in some sense, serves as the object's structural backbone. The trunk of a tree or the fuselage of an airplane, for instance, are the most well-connected parts within the objects. It is conceivable such parts are special because they are the most stable (e.g., the trunk of the tree is the part least likely to bend with the wind) or because they carry some functional significance. Alternatively, the elongated part may be important simply because they are the most prominent in size and this saliency tends to draw the observer's attention.

CHAPTER 6. GENERAL DISCUSSION

The current thesis aimed to achieve two goals. First, I probed the mechanism for representing the spatial relationships of parts within an object. The first two experiments hypothesized that if the Coordinate Orientation Representation (COR) theory could be adapted to account for the way object parts' relations were represented, common errors in the shape recall task would involve reflections of parts within an object. In Experiment 1, I found that when participants made an orientation error in recalling the internal configuration of an artificial object, their error frequently corresponded to a reflection of an object part across its intrinsic principal axis (i.e., IPA error) and across the intrinsic secondary axis (i.e., ISA error). The existence of these errors was not easily explained by alternative views of shape or axis representation, but were consistent with the COR-based account that object parts' relations were represented based on systems of coordinate axes (see § 3.6.1. Representation of Relative Part Orientation).

The COR-based account also predicted that errors in the form of location reflection would be common in the shape recall task (see § 3.6.2. Representation of Relative Part Location). The first experiment, however, was not ideal for probing this type of errors because the way that the way the stimuli were designed did not permit some types of location reflection to occur. In Experiment 2, I eliminated this problem by modifying the stimulus objects to make all types of location reflection equally plausible. Participants in this experiment indeed tended to commit location reflection errors, further supporting the notion that the COR theory could be adapted to explain how object parts' locations were represented. Besides the main finding, the first two experiments revealed that when participants misremembered how an object part tilted, their errors revealed interesting insights about how the object part's location was defined. I found that the basis for defining a part's location changes with modifications in stimulus design, and discussed this finding in terms of possible changes in the underlying representation.

Experiment 3 and 4 pursued the second goal of the thesis. In particular, I asked whether the origin of shape coordinate axes (i.e., the point where the principal and the secondary axes meet) is computed on the basis of an individual object part (i.e., the Elongated-Center Hypothesis), or by taking all parts or regions of an object into account in some way (i.e., the Object-Center Hypothesis). The result of Experiment 4 supported the Object-Center Hypothesis. I discussed the result by contrasting it with the previous finding about how axes' tilts were defined (Chaisilprungraung et al., 2019), and offered explanations in terms of potential advantages of different ways of defining the axes.

6.1. Specific Mechanism for Processing Shapes?

A critical proposal of this thesis is that the mechanism for representing object parts is similar to the mechanism for representing whole objects, in that both require relating coordinate axes across frames of reference after the manner posited by the COR theory. In the shape recall experiments, was it possible that what was probed was not a mechanism for representing object parts, but a mechanism for processing whole objects? If participants treated the parts of an artificial stimulus as two independent shapes rather than components from within the same shape, they might attempt to carry out the task by remembering where an object is (or how it is oriented) with respect to an extrinsic frame of reference. Could such a strategy be possible, or could it produce the error patterns observed in the current thesis?

It is important to remember that the shape recall task was designed so that the most effective way to succeed was to remember how object parts were internally related (see §5.2. Shape Recall Task). In each trial a target stimulus was displayed at three random locations and orientations on the screen. After the target disappeared the probe stimulus was presented at a new location and orientation, and participants had to modify the probe so that its internal configuration matched the target's. Because the relationship between the stimulus and an extrinsic frame of reference (e.g., the screen) was always changing, the most natural way to perform the task was by focusing on how the stimulus' parts were internally related, and ignore any possible relationship between the stimulus (or parts of the stimulus) and the extrinsic reference frame.

If participants treated the artificial stimulus as composed of two independent shapes (i.e., a large shape and a small shape), they had to somehow transform each individual shape so that the shapes were in the right relation with respect to one another. Such strategies were not impossible. For instance, during the target presentation they might chose to pay attention to only the final presentation and ignore the first two. When the probe appeared after the last presentation, they might try to transform the large and the small shapes individually, so that the orientation of each shape with respect to the screen exactly matched what they saw in the final presentation.[50] In a further analysis we identified the number of responses where the absolute orientations of both the large and the small shapes were within 30˚ (in either clockwise or counterclockwise direction) of the correct answer in any of the three presentations. We found that such responses accounted for a small fraction of the current data[51], indicating that participants did not

[50] Since participants were not allowed to move the large object part, only the orientation, and not the location of the shapes could be matched.
[51] 5.7% of the total data in Experiment 1 (i.e., 133 of 2304 trials) and 9.9% of the data in Experiment 2 (i.e., 214 or 2160 trials).

primarily rely on this strategy to solve the task.

Another possible strategy may involve piecemeal mental transformations of the large and the small shapes. For example, by imagining how the large shape must have been rotated in order to produce a new orientation, participants could have tried to solve the task by applying the same mental rotation to the small shape. An important challenge for this interpretation is that it is unclear how it can account for the different error patterns observed across experiments. In Experiment 2, we found that errors in the form of 70˚ rotation occurred at a very high rate in addition to the IPA and the ISA errors. In addition, the relative location preserved in the small shape's responses corresponded to the place of attachment rather than the shape's center. Under the COR-based theory of shape representation, it was possible to attribute such errors to critical modifications in the stimulus design (i.e., the addition of a joint features, the change in the relative tilt of object parts in the presentation), and possible changes in parameters of the representation that resulted from these modifications (see discussion in §5.3.3). Under the interpretation where the parts were independently transformed, however, there was no clear reason why the modifications should have produced changes in the error pattern. If separate mental transformations were applied to the large and the small shapes, then changing the angular degree at which the shapes were tilted with respect to one another, or adding a feature indicating how the shapes were connected should have had no effect on the error pattern.

Perhaps a fruitful question that should be asked related to this topic is what property(s) of the parts of the artificial stimuli made them 'part-like', instead of 'object-like'. In the shape recall ask the large and the small object part were always displayed as physically joined to one another. Might physical attachment be as an important criterion for whether two shapes are treated as being parts of the same object? To test this hypothesis a future study may be carried out where the parts are displayed nearby one another on the screen but do not directly touch one another. If physical attachment is critical for establishing a part status, then separating the large and the small components of the artificial object should produce a different error pattern. Alternatively, the important criterion may be whether the spatial relationships between the parts are maintained across presentations. Parts of real-world objects often move or rotate with one another when acted upon. When a teapot is lifted from a tray and tilted over a teacup, the internal relationships among the spout, the lid, and the handle are maintained, but the external relationships change (e.g., relationships between the spout and the teacup or the tray). If the constancy of relationships is what matters in shape representation, we may expect a different pattern of result when the constancy is disrupted. In a modified shape recall experiment, the procedure may be designed so that in some random trials, participants see the relations between the large and the small object parts changed across the three presentations, and are asked to do something else besides recalling the parts' relations (e.g., recalling where each stimulus is presented on the screen). Alternatively, it could be that what is important is the relative size of the object parts. If the small object part was made larger or so that it had the equal size as the large object part, for example, might they be more likely to be treated as independent objects rather than parts from within the same shape? These questions merit further investigation.

6.2. Levels of Representation

Another important question that should be asked is what level(s) of visual representation were probed in the experiments. Some researchers posited that axis-based representations are computed prior to and provide the basis for object recognition (e.g., Large et al., 2003; Leek & Johnston, 2006; Ling & Sanocki, 1995; Marr & Nishihara, 1978; Quinlan and Humphreys, 1993; Sekuler & Swimmer, 2000). For the current thesis, could the reflection errors occur because of some failures in the way axes are computed before the recognition stage?

There are a few reasons to suggest that this is unlikely. Across all four experiments participants were asked to recall the shape or the orientation of a target stimulus after a retention interval of several seconds. During this interval, the participants also carried out additional visual shape processing (e.g., viewing or recalling the second stimulus object). Visual recognition however is usually achieved in less than a second (Potter, 1976; Thorpe et al., 1996). Thus due to the way the experiments were designed, it was more likely that the representations directly mediating participants' responses were computed after the object was recognized.

Might it still be possible that the errors produced during postrecognition processing were actually introduced during a prerecognition stage? A representational failure affecting an early stage which was preserved through later levels of processing can still be manifested in the eventual response. Could it be that common errors like intrinsic-axis reflections (e.g., IPA and the ISA errors) occurred because of failures affecting polarity correspondence when the object was first encoded? We have informal data to suggest that this is unlikely. Prior to conducting Experiment 4 we conducted a pilot study where one object was tested per trial instead of two objects. Results from this study yielded highly accurate performance, where the rate of reflection errors was only 3.3% (i.e., approximately two errors per 60 responses) compared to the 18.3% (i.e., approximately

11 errors) obtained from the actual study.[52] The result therefore challenges the possibility that common reflection errors were introduced during the encoding stage of shape processing.

6.3. Axes for Everything?

The current discussion about coordinate axes was largely conducted in the context of shapes with clearly identifiable parts (e.g., the large vs. the small part of the novel shape). How might the notion of axes apply to objects or shapes without a clear part (e.g., a crumbled ball of newspaper, a blob)? To that end, are there axes for substances or materials that lack a solid form (e.g., water, sand, cloud), invisible energies (e.g., sound, light, electricity, forces), or non-concrete things (e.g., color, smell, time, idea)?

Perhaps the best way to address these questions is to phrase the questions in two slightly different ways. First, it can be asked whether it is theoretically possible to define axes to something that lack clear parts or that to which the notion of parts does not apply. Second, assuming that it is possible to define the axes, it can be asked whether such axes are useful or relevant to a perceptual task--that is, whether the observers actually compute or use the axes in processing spatial information. I address each question below.

Possibility of defining an axis

The answer to whether it is possible to define an axis depends on what is meant by an 'axis'. In the current thesis, axes are construed as something that characterize spatial directions within a reference frame. The axes of a table may characterize a direction in which the table is elongated and the direction where the legs extend from the tabletop.

---

[52] Indeed, the result from the pilot study motivated our decision to use two-object-per-trial design to ensure that we obtained enough reflection data.

Based on this construal, it is possible to define an axis (or axes) so long as it is possible to conceive of spatial directions. All solid shapes that are not perfectly symmetrical (i.e., that are not perfectly spherical or circular) can have an axis. This is true regardless of how complex the shape is (e.g., a tangle of wire) or whether the shape has clearly identifiable parts (e.g., a ball of crumbled newspaper, a blob). A blob, for instance, can theoretically have an axis if it is elongated or if it can be partitioned into asymmetrical halves.[53] For substances without a solid form or which are invisible (i.e., water, gases, electricity, forces), axes may be present if spatial directions can be defined based on other intrinsic properties. The axis of water in a river or a stream, for example, can be defined based on the direction of the water flow. Similarly, the axis of the air can be defined based on the direction of the wind, and the axis of the sky based on differing amount of light, oxygen, temperature, or pressure that exist across different parts of the atmosphere (e.g., stratosphere vs. ionosphere). Invisible forces and energies often possess an axis that can be defined based on the force's direction (e.g., the direction of gravity, the direction of magnetic or frictional force [54]), or the direction travelled by waves or small particles (e.g., the direction that sound or light travel, the direction of electric current).

If it is not possible to define spatial directions, however, then axes cannot be present. Still water in a glass or air in a balloon cannot have an axis.[55] Similarly abstract

[53] Indeed, elongated or asymmetrical blobs are common stimuli in psychological studies of shape axes (e.g., Cohen & Singh, 2006; Morikawa, 1999).

[54] Indeed, the axis defined by the direction of gravity is central to some theories of shape representation (e.g., Attneave & Olson, 1967; Rock, 1973; see §2.2), and the axis based the earth's magnetic field provides an important basis for defining map notions such as 'north' and 'south'.

[55] Although, the possibility of defining axes for fluids in a container may depend on whether it is possible to assume that the fluids can have intrinsic shapes independent of the container. For example, does water in a tube have an elongated shape? If so, the water can have an axis of elongation.

things like time, color, smell, thoughts or ideas do not have intrinsic spatial directions and therefore, they cannot possess an axis.[56]

Relevance to cognition

The important question, however, is not whether axes can theoretically be defined but whether they are relevant to the way observers process spatial information. Although it is not impossible to define axes for a blob or a crumbled ball of newspaper, it is difficult to imagine how such axes are helpful in the way the shapes are mentally represented. Since these shapes lack easily identifiable parts, it may be reasonable to argue that axes are not perceptually defined since the shapes are not processed in terms of parts. However, even for these shapes there must be some way of representing the overall orientations. Under the assumption where objects' orientations are represented as relationships between reference frames, coordinate axes must be available for the object's frame of reference. Since these shapes lacks clear axes of elongation or axes of symmetry, however, it is conceivable that multiple ways of defining the axes are acceptable.

It is important to observe that the likelihood that something is processed in terms of parts and relations between parts also depend on the specific requirements of the task at hand. An observer who intends to throw a crumbled newspaper into a wastebasket may not care about the object's spatial details, and therefore does not analyze the object in terms of parts. However, an observer who tries to create a sketch of the crumbled newspaper may be forced to carefully study the object's spatial details, and thereby may

---

[56] It may be possible to define other types of axes for things that lack spatial directions. For instance, time has a temporal axis that can be defined based on the relative past vs. present. Abstract ideas or thoughts can have axes defined based on opposing concepts (e.g., sadness vs. happiness, justice vs. corruption).

analyze the shape in terms of component parts and parts' relations. Even for objects with clearly identifiable parts, task requirements may determine whether information about object parts' relations, and therefore object axes, is used in object processing. An observer searching for an object on a supermarket shelf full of similar-looking items (e.g., a ketchup among different sauce bottles) may find other salient visual features (e.g., color, size) more efficient for the search. In such a situation information about how object parts are related may not be used even if computed.

Similarly, although it is difficult to imagine how axes of invisible or non-solid things are relevant to daily life visual processing, the usefulness of axes may depend on context and circumstances. For sailors, pilots, and different species of bird and fish, axes defined by the wind's direction or the direction of waterflow are key to spatial navigation. Some creatures, including desert ants, use polarization of light in the sky to define directions for navigation. In addition, some blind individuals and bats and whales can rely on direction of sound in learning the shape of the environment and wayfinding.

In short, it is possible to define axes so long as it is possible to conceive of spatial directions. The relevance and usefulness of axes in spatial cognition, however, may critically depend on the context and the requirements of the task at hand. In the current thesis, information about how object parts are related is clearly important for success in the shape recall task. Future research may benefit from exploring whether shape axes are still used in a task where information about how object parts are related is less important.

6.4. Conclusion

The question of how the brain perceives and represents spatial structures of objects constitutes a central topic of psychological research on vision. Previous studies assumed that the spatial structure of an object is represented with respect to some axes defined based on some structural properties of shapes. But it is unclear what mechanism(s) is precisely involved in encoding relationships of object parts with respect to the axes. In this thesis, I consider the possibility that a mechanism posited by the COR theory, which was originally proposed to account for how the brain represents orientations of whole objects in an environment, can be adapted to explain how object part relationships are defined. The first two experiments provided evidence in support of this possibility, suggesting that relations of object parts were represented based on systems of coordinate axes. Another important question examined in the thesis is how the origin of the coordinate axes is defined. The latter two experiments found that axes' origins are defined based on the center of the overall object, thereby increasing the current understanding of shape axes.

The current thesis has several potential contributions to research on object shape processing. First, it highlights the importance of taking into account spatial directions and internal part relations when theorizing about shape representation. Second, it presented a series of novel findings suggesting that a theory about whole objects' orientations could be adapted to explain how object parts' relations were represented. Third, it introduces a novel experimental paradigm where insights about object shapes could be obtained based on analyses of mirror-image reflection errors. Finally, it provides a novel empirical ground for which to compare different accounts of shape processing including those that are inspired from computer vision research (e.g., Medial Axes, Convolutional Neural Network).

APPENDIX A. Robustness of Reflection Pattern

A.1. Experiment 1

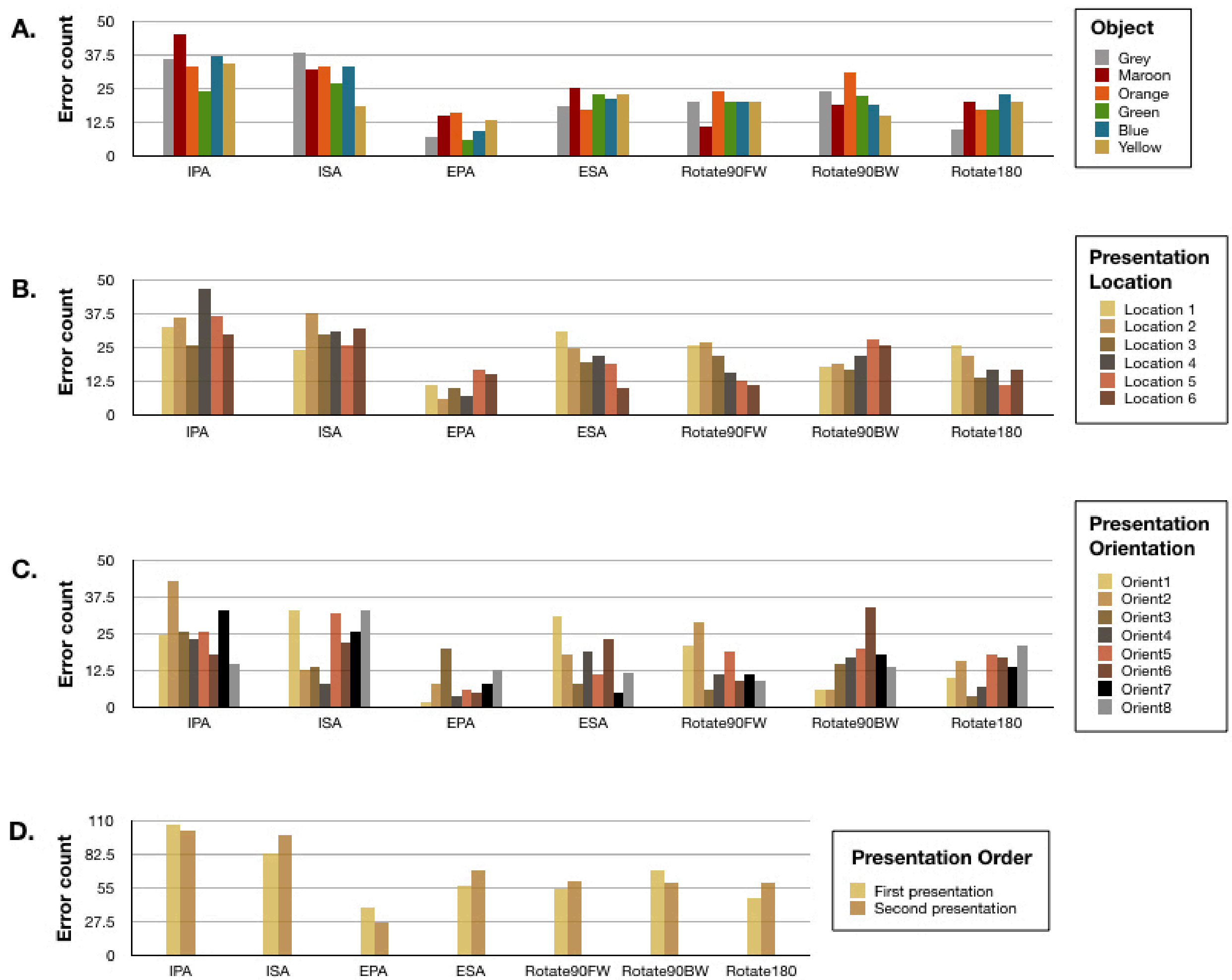

Figure A- i *Plots showing the distribution of orientation error in Experiment 1. The distribution is displayed across (A) different types of stimulus object, (B&C) different locations and orientations of target presentation, and (D) different orders of target presentation within a trial.*

A.2. Experiment 2

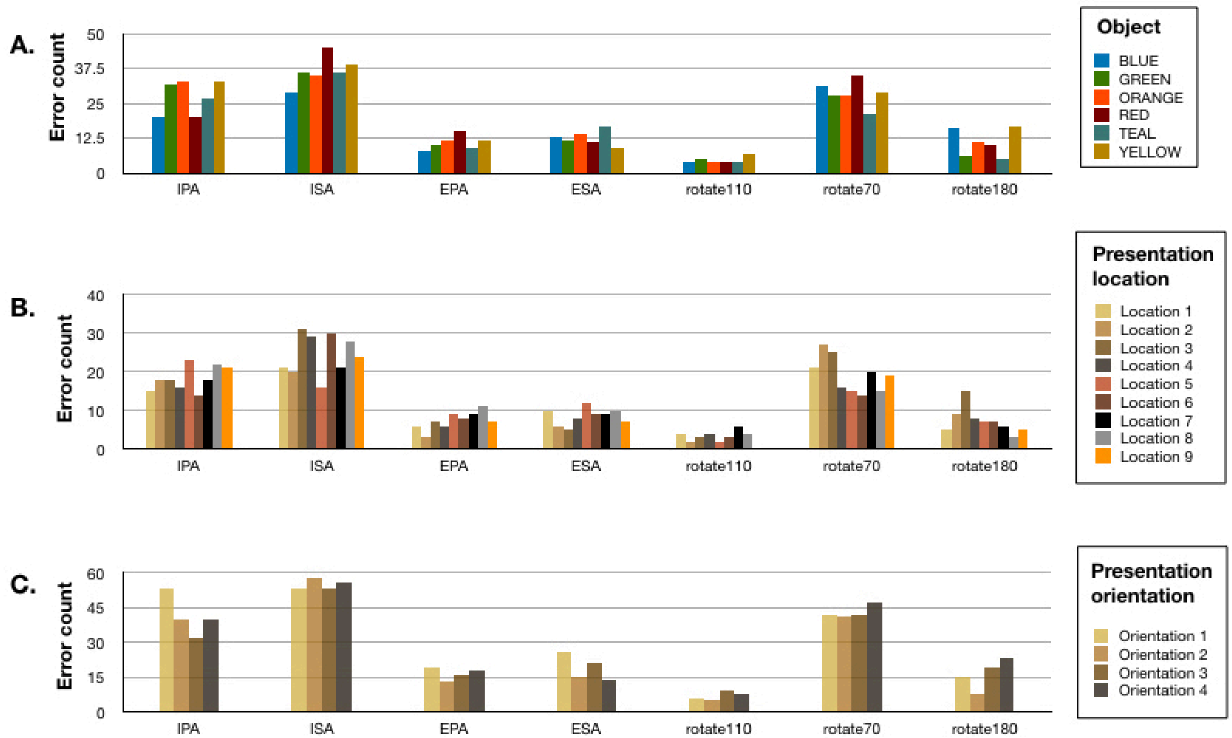

Figure A- ii. *Plots showing the distribution of orientation error in Experiment 2. The distribution is displayed across (A) different types of stimulus object and (B&C) different locations and orientations of target presentation.*

A.3. Experiment 3

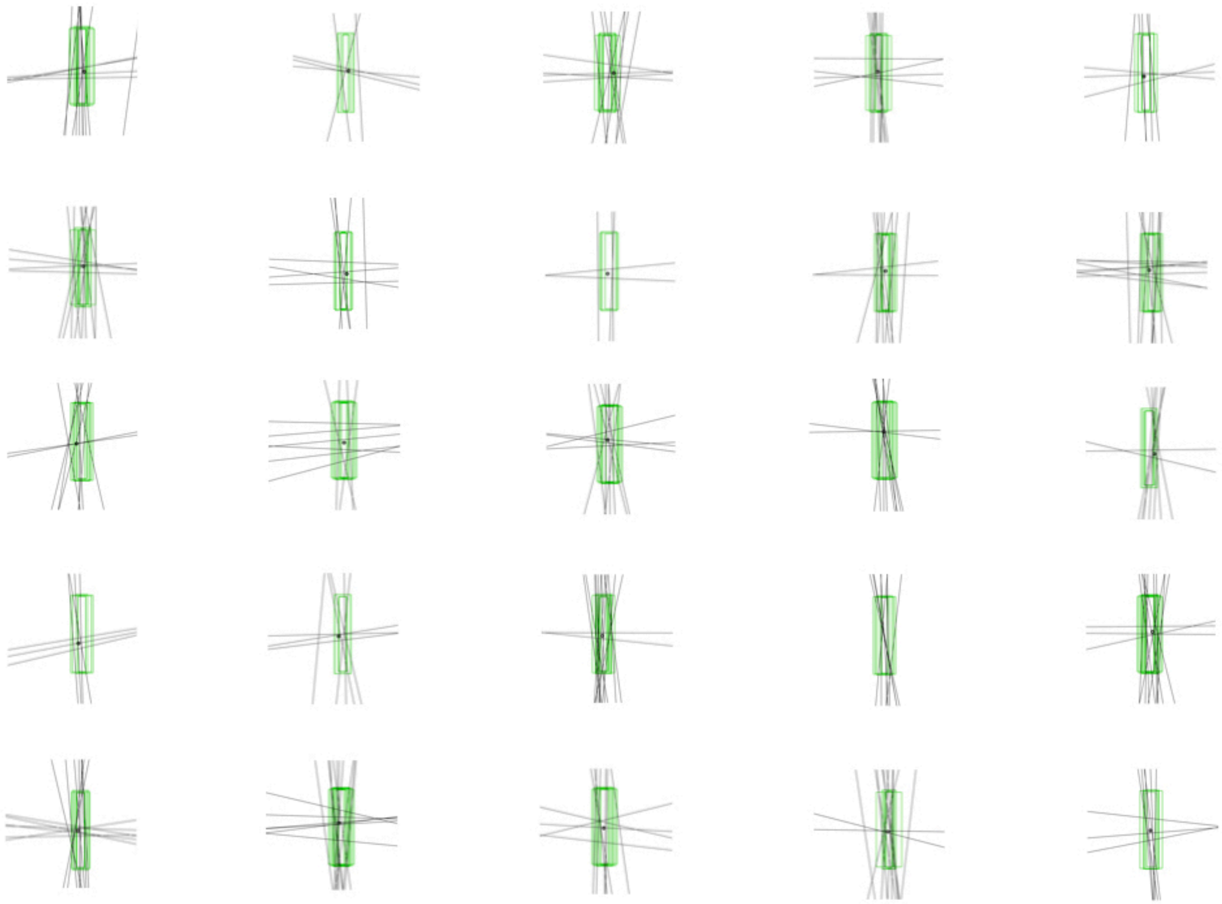

Figure A- iii. *Plots showing axes of reflection and the average point of axis intersection for all 25 presentation locations in Experiment 3. The plots' spatial arrangement corresponds to the actual 5x5 locations that the target could be presented on the screen.*

A.4. Experiment 4

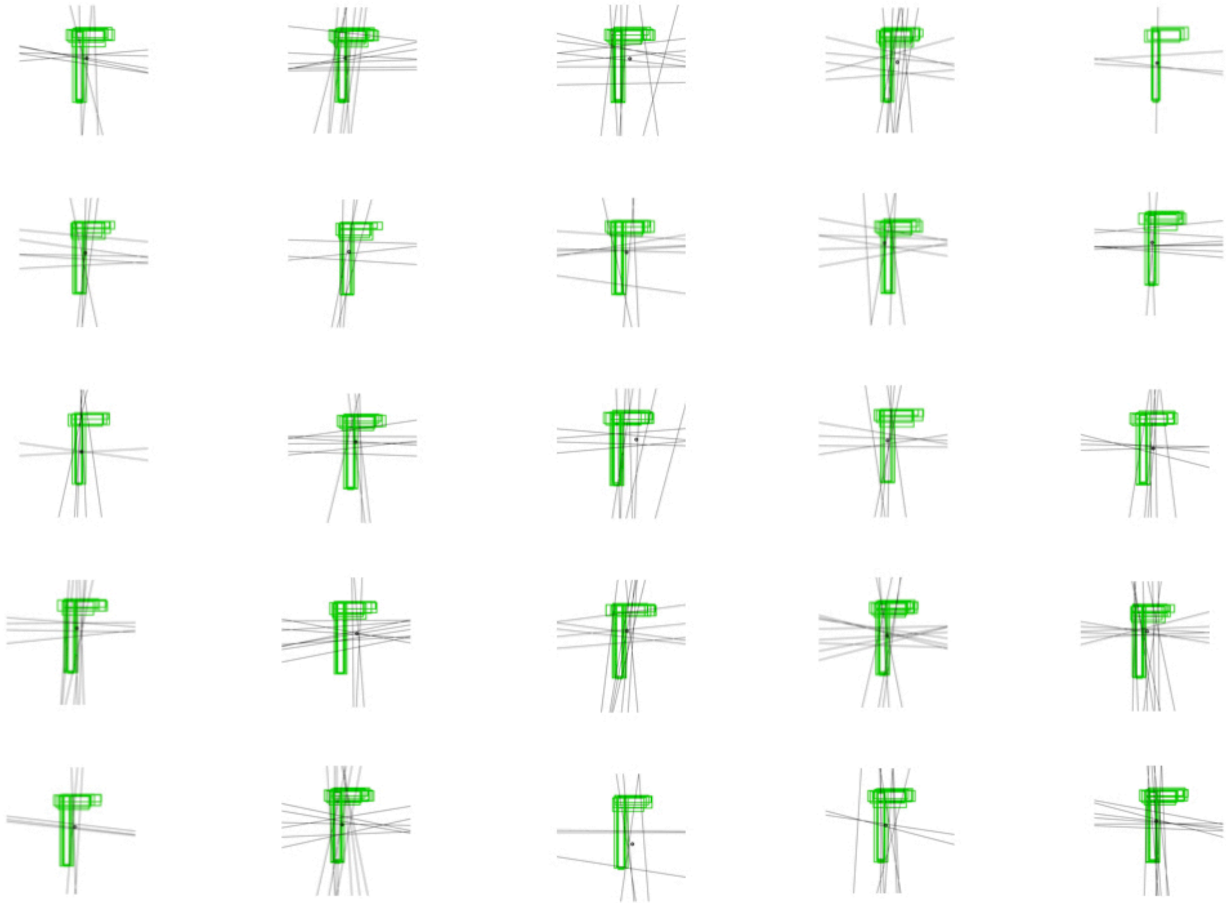

Figure A- iv. *Plots showing axes of reflection and the average point of axis intersection for all 25 presentation locations in Experiment 4. The plots' spatial arrangement corresponds to the actual 5x5 locations that the target could be presented on the screen.*

APPENDIX B. Convolutional Neural Networks Analysis

B.1. Experiment 1

The CNN analysis was performed by testing the pre-trained Alex Net (Krizhevsky et al., 2012) on a set of stimulus images that participants saw in the shape recall task. These images corresponded to all possible combinations of six stimulus objects, 48 part configurations, and three whole-object orientations (see Figure B- i below). Each image was rescaled to the resolution size of 277 x 277 pixels as required by Alex Net, and was then processed through multiple layers of the model.

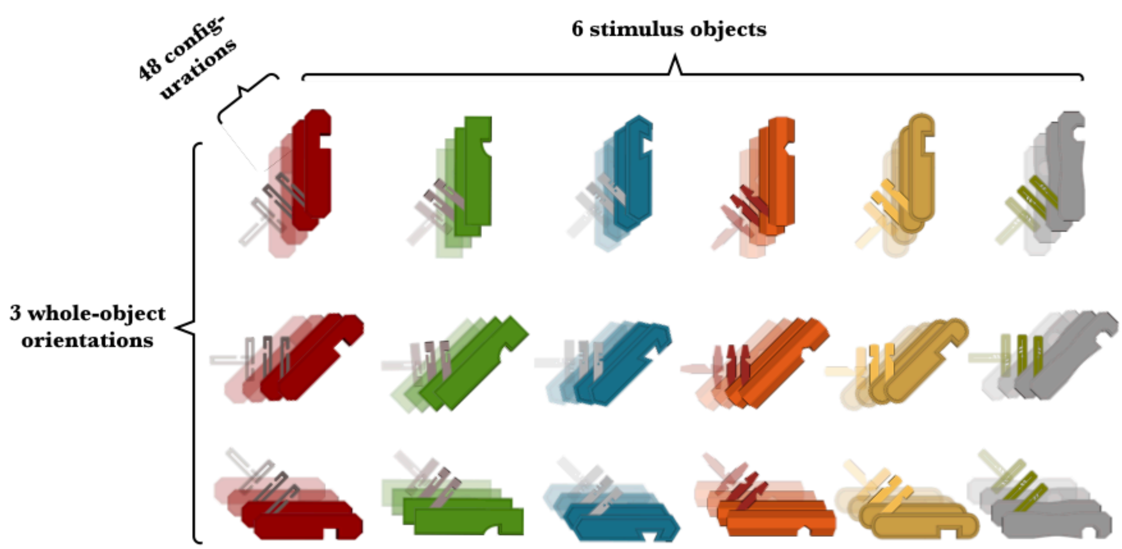

Figure B- i *Image inputs to Alex Net*

Figure B- ii illustrates the organization of layers in Alex Net. The first layer takes the rescaled image, which corresponds to a matrix of 227x227x3 pixel values, and passes a series of filters (also called 'kernels') over the image. These filters are small matrices of a particular size (11x11 for the first convolution layer) (see Figure B- ii). Depending on the values inside the filters (which were set through prior learning), different features of the image are extracted (e.g., edges, curves). In the first convolution layer of Alex Net, 96 different filters are used, producing a 55x55x96 matrix containing transformed pixel values (i.e., the so-called 'activation map' or 'feature map'). After convolution, an additional operation is performed in the first layer which replaces all negative pixel values in the activation map with zero. This process (i.e., the so-called 'ReLU' or Rectified Linear Unit operation) has the effect of introducing non-linearity to CNN, which is assumed to be necessary because most real-world data is non-linear.

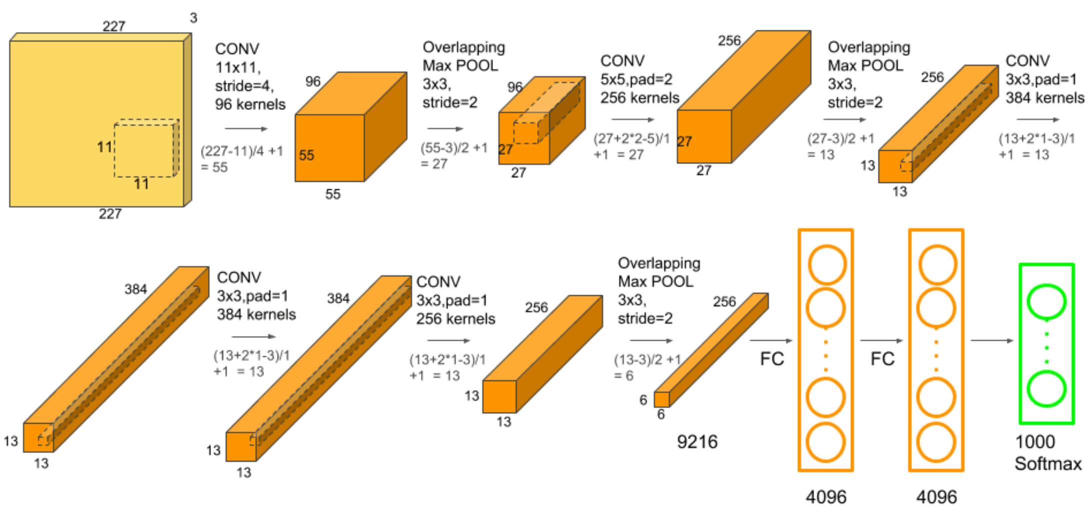

***Figure B- ii** Architecture of Alex Net*

The first convolution layer is followed by the second layer involving image pooling. In pooling, the model divides the activation map into small square windows (e.g., 3x3 windows for Alex Net), and within each window, retains only the value that satisfies a certain criterion. In CNN, only the maximum values are retained. The max pooling step reduces the size of the activation map, making the map more manageable for later processing. In addition, image pooling is assumed to be important in that it makes the network invariant to small distortions and transformations (e.g., scaling, translating) of the visual input.

The first two layers involving image convolution and pooling are repeated, and followed by three convolution layers and another pooling step (Figure B- ii). Across the five convolution layers in the model, the number and the size of the filters are not the same [57], and these parameters were set in advance so that they do not get updated during learning.

In the final three layers, each unit in the activation map is directly connected across layers in the one-to-one fashion. The fully-connected (FC) layers take the output from the convolution and the pooling layers which is assumed to contain high-level representation of the image, and convert them to a vector of 1000 values, each containing the probability that the image belongs to a particular category.

After the images were passed through all the layers, I extracted the values inside the activation map in each layer, and performed a series of pairwise correlations (Pearson's r) or each possible of the image inputs (Figure B- i). The pairwise correlations for each layer of CNN are displayed in the plot in the Discussion section of Chapter 5 (Figure 5-10C). To examine whether the correlation pattern was robust, I further analyzed the pairwise correlations separately for each type of stimulus object, each whole-object orientation, and each internal configuration of object part (i.e., relative part orientation and location). The results, shown in Figure B- iii below, suggest that the correlation pattern is highly robust.

[57] across five convolution layers, the numbers of filters are 96, 256, 384, 384, 256, and size 11, 5, 3, 3, 3.

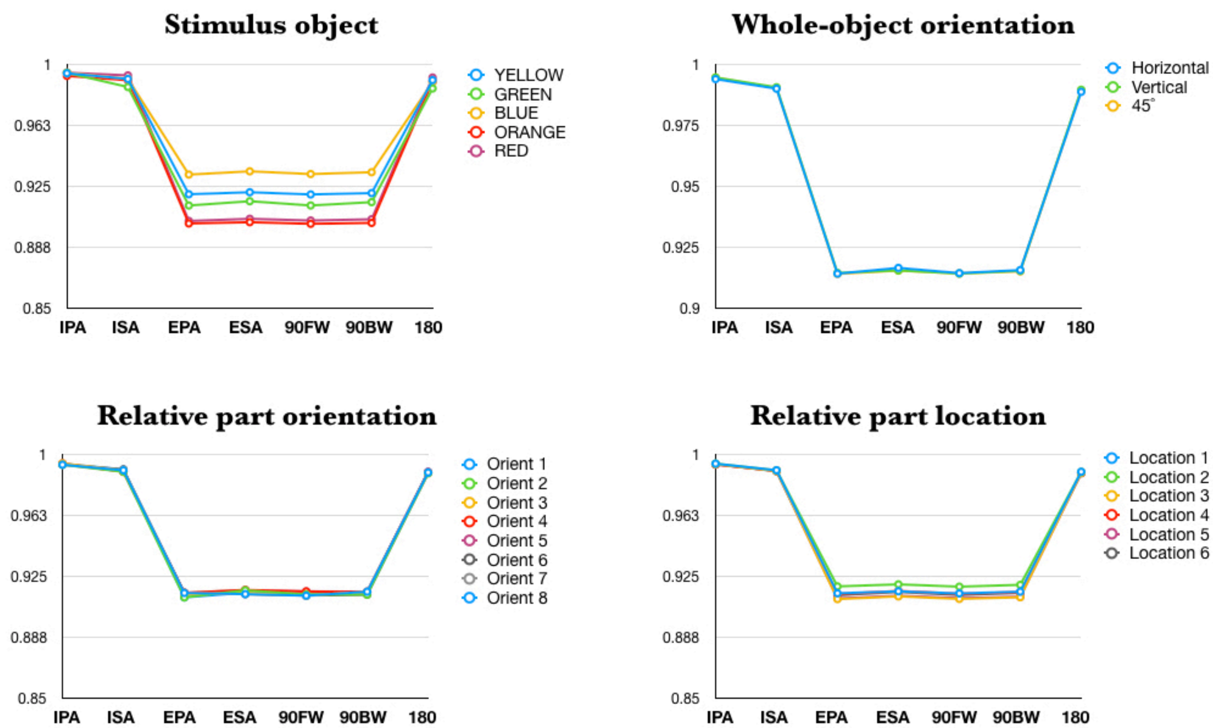

Figure B- iii *Charts displaying CNN results across conditions of target presentation. Y-Axis represents the correlation of CNN activation (Pearson's r) between an error type (X-axis) and the corresponding target.*

B.2. Experiment 2

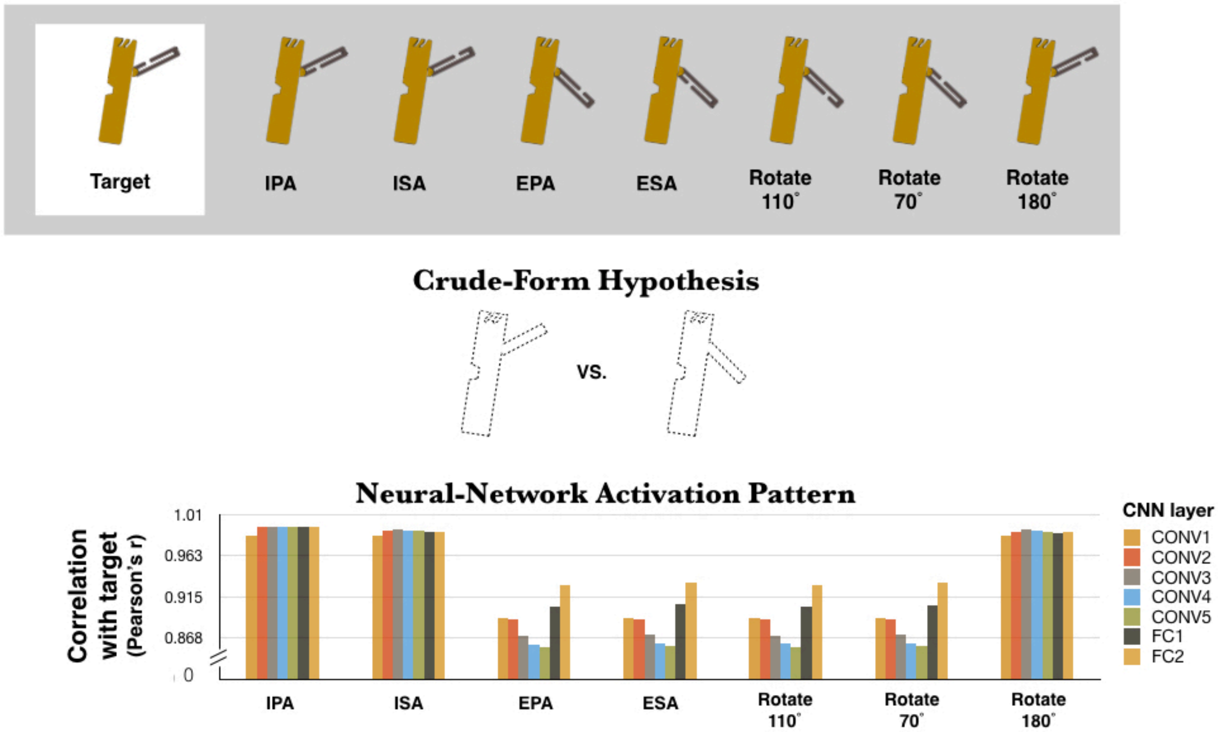

Figure B- iv *Illustration of the crude forms of the stimulus objects, and the results of the convolution neural network (CNN) analysis. Error bars denote 95% CI.*

APPENDIX C. Tilts of Reflection Axes

C.1. Experiment 3

To examine the tilts of the reflection axes, we created a plot similar to the previous orientation recall study (Chaisilprungraung et al., 2019), which exclusively displays the orientations of the reflection axes (Figure C- i). In this plot the axes' orientations are defined relative to the stimulus object's axis of elongation. Reflection-axis orientations of approximately 0° correspond to reflections across the object's axis of elongation, and reflection axes of approximately 90° correspond to reflections across an axis perpendicular to the object's elongation.

We computed the pair of axes that best fit the reflection-axis data—more specifically, the perpendicular axes that minimized the sum of absolute angular distances to the individual reflection axes. These best-fit axes had orientations of -1˚ and 89˚ (see tick marks in the figure), which were very close to the orientations of the posited principal and secondary object axes. The result suggested that the tilt pattern of the reflection axes replicated the previous finding (Chaisilprungraung et al., 2019, see Figure 3-9).

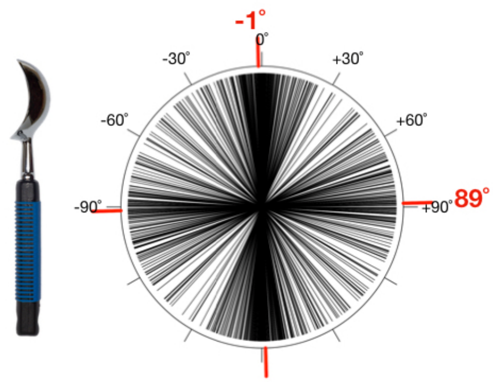

Figure C- i. *Reflection axes of Experiment 3 displayed in a circular plot. Large tick marks designate the pair of perpendicular axes that best fit the reflection data*

C.2. Experiment 4

A similar analysis was performed to examine the tilts of the reflection axes of Experiment 4. The analysis revealed the best-fit orientations of 5˚ and 85˚ (see tick marks in the figure below), which somewhat differed from the 1˚ and -89˚ orientations observed in the previous study (Chaisilprungraung et al., 2019, see Figure 4-4C). The difference in the best-fit orientations might be due to the presence of some reflection axes that corresponded neither to the principal nor the secondary axis of the object. Specifically, we observed some axes shown as the left-diagonal and the right-diagonal clusters in the circular plot. These axes might correspond to noise that occurred due to the increased task difficulty, or they might correspond to reflection errors across the object's axis of symmetry (i.e., the axes that bisected the junction between the elongated part and the protruding part), and the axis perpendicular with the symmetry. The presence of the axes could affect the reliability of the best-fit computation, which was performed under the assumption that only two orthogonal clusters of axes were present[58].

[58] Indeed, our informal simulation analysis confirmed that the best-fit orientations were highly unreliable when the distribution of the reflection axes was not bi-modal.

This intuition was supported by results from a further analysis: When the reflection data was separated according to the order of stimulus presentation within a trial (see the two bottom figures), different patterns of reflection axes were observed. The stimuli from the first presentation were typically the easier ones to recall, and the axis pattern clearly suggested the frequent occurrence of reflections across object axes defined based on the elongated part. The best-fit orientations of these stimuli were also closer to what was observed in the previous study (i.e., 3.5˚/-86.5˚).

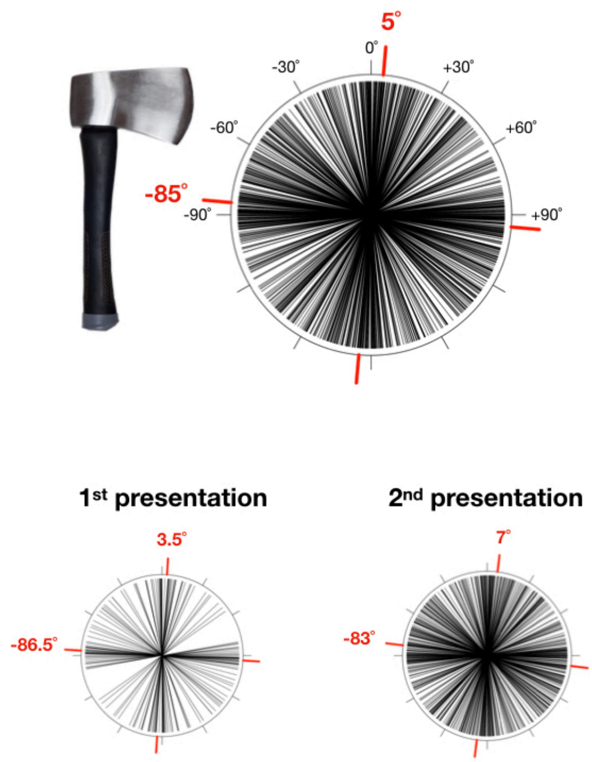

Figure C- ii. *Reflection axes of Experiment 4 displayed in a circular plot. The bottom plots of the figure show reflection axes obtained for the first (left) and the second stimulus presentation in a trial. Large tick marks designate the pair of perpendicular axes that best fit the reflection data*

APPENDIX D. Analysis of Reflection Data Simulation

We first created simulated responses of participants that closely resembled the data collected from Experiment 3 as much as possible. Descriptive coefficients of participants' errors like the rates of reflection, the means and the standard deviations of the tilt and the location errors were extracted from the original results (see first row of Table D- i below), and were used as bases for constructing the simulated data. Like the actual experiment, the total of 1,800 simulated responses were created. The tilt errors (i.e., the angular deviations from target) of the simulated responses were randomly sampled from a normal distribution with mean at 0˚ and the standard deviation at 15˚. The location displacements of the responses were randomly sampled separately for the X- and the Y-axis, both from a normal distribution with mean at 0 and standard deviation 0.06. From these simulated responses, 278 were selected at random to have an incorrect enantiomorph--193 were selected to correspond to reflections across the object principal axis (i.e., OPA errors), and 85 to reflections across the object secondary axis (i.e., OSA errors)[59].

Based on the simulated responses, we performed an analysis to identify the axes of reflection and the points of intersection, using the same analysis method described in Experiment 3 (see 5.4.2. Experiment 3 Results). The results, shown in Figure D- i below, reveal axes of reflection with the red circle denoting the point of intersection averaged across different types of objects. The mean location was 0.00069 along the X-axis and 0.00015 along the Y-axis. Single-sample t-test revealed that this mean location was not

[59] To create the OSA errors, we also had to rotate the responses 180˚, in addition to changing the enantiomorphs.

significantly different from the object's center (i.e., $t(11) = -0.003, p = 0.9$ for location along X-axis and $t(11) = 0.23, p = 0.8$ for location along Y-axis). The results suggested that the current simulation method can be used to model the reflection data.

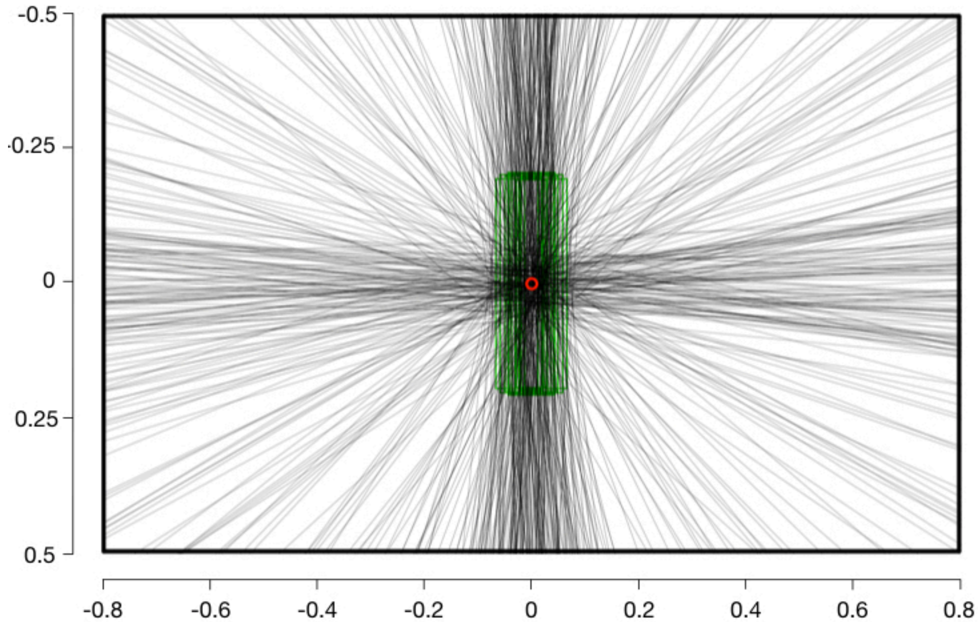

Figure D- i. *Axes of reflection computed from the simulated responses of participants. The red circle denotes the average points of intersection.*

Based on the success of the above simulation, we created three sets of simulated data based on modified coefficients of participants' errors. Table D- i shows the coefficients that were used as bases for each data simulation. The red texts highlight the coefficient that was changed in each simulation. In the first simulation (see second row of the table), the mean tilt of participants' errors was altered from 0˚ to 15˚. The result of this modification was equivalent to assuming that participants frequently remembered the stimuli as being tilted 15˚ in the clockwise direction relative to the orientations that they were originally displayed. In the second simulation (third row of the table), the standard deviation of the location errors was increased from 0.06 to 0.12, which was equivalent to assuming that participants were on average twice as inaccurate in remembering the stimuli's location as was originally observed. In the third simulation (final row of the table), the mean of the location errors was changed from (0,0) to (-0.05,0.1). This was equivalent to assuming that participants tended to remember the target stimuli as being closer to the bottom left of the screen.

	Number of reflection errors		Tilt error	Location error
	OPA	OSA		
Simulated data similar to Experiment 1	193	85	Mean = 0° SD = 15°	$Mean_{(x,y)}$ = (0,0) $SD_{(x,y)}$ = (0.06,0.06)
A. Simulated data with bias in tilt direction	193	85	Mean = 15° SD = 15°	$Mean_{(x,y)}$ = (0,0) $SD_{(x,y)}$ = (0.06,0.06)
B. Simulated data with increased amount of location error	193	85	Mean = 0° SD = 15°	$Mean_{(x,y)}$ = (0,0) $SD_{(x,y)}$ = (0.12,0.12)
C. Simulated data with bias in location direction	193	85	Mean = 0° SD = 15°	$Mean_{(x,y)}$ = (-0.05,0.1) $SD_{(x,y)}$ = (0.06,0.06)

Table D- i. *Table showing coefficients that were used as bases for creating simulated responses of participants.*

The left panel of Figure D- ii shows plots of the tilt and location errors, which illustrate changes in distribution that occur in each simulation. The right panel of the figure shows axes of reflection computed from the simulated data. As the results show, the average point of axis intersection remains at the center of the elongated object, regardless of how the error distribution was manipulated. Changing the mean of the tilt error (Figure D- iiA) also affect the tilts of the reflection axes, but the common points of axis intersection still correspond to the object's center. Similar conclusions apply to the other two simulations (Figure D- iiB&C).

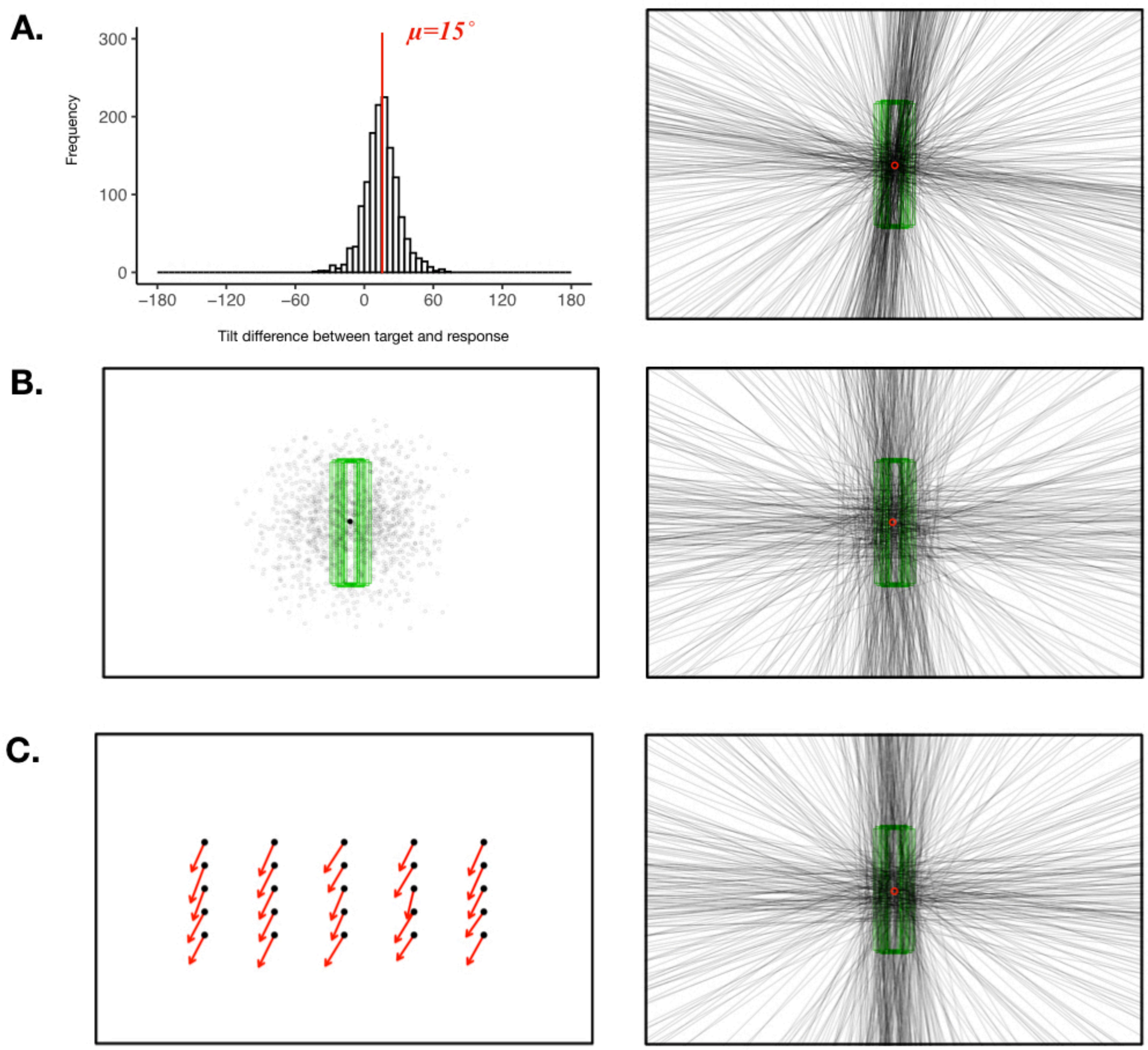

Figure D- ii. *Plots illustrating how the tilt and the location errors were changed in each data simulation (left panel), and the corresponding results (right panel).*

APPENDIX E. Manually Drawn Bounding Boxes

Figure E- i. *Illustration showing bounding boxes that were manually drawn for parts within an object.*

BIBILOGRAPHY

Attneave, F., & Olson, R. K. (1967). Discriminability of stimuli varying in physical and retinal orientation. *Journal of Experimental Psychology*, *74*(2p1), 149.

Ayzenberg, V., Chen, Y., Yousif, S. R., & Lourenco, S. F. (2019). Skeletal representations of shape in human vision: Evidence for a pruned medial axis model. *Journal of Vision*, *19*(6), 6–6.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *1*(4), 371–394.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115.

Blum, H., & Nagel, R. N. (1978). Shape description using weighted symmetric axis features. *Pattern Recognition*, *10*(3), 167–180.

Boutsen, L., & Marendaz, C. (2001). Detection of shape orientation depends on salient axes of symmetry and elongation: Evidence from visual search. *Perception & Psychophysics*, *63*(3), 404–422.

Bradshaw, G., & O'Sullivan, C. (2004). Adaptive medial-axis approximation for sphere-tree construction. *ACM Transactions on Graphics (TOG)*, *23*(1), 1–26.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Chaisilprungraung, T., German, J., & McCloskey, M. (2019). How are object shape axes defined? Evidence from mirror-image confusions. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(1), 111.

Cohen, E. H., & Singh, M. (2006). Perceived orientation of complex shape reflects graded part decomposition. *Journal of Vision*, *6*(8), 4–4.

Connell, J. H. (1985). *Learning Shape Descriptions: Generating and Generalizing Models of Visual Objects.* MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB.

Cornea, N. D., Silver, D., & Min, P. (2007). Curve-skeleton properties, applications, and algorithms. *IEEE Transactions on Visualization and Computer Graphics*, *13*(3), 0530–548.

Davidoff, J., & Warrington, E. K. (2001). A particular difficulty in discriminating between mirror images. *Neuropsychologia*, *39*(10), 1022–1036.

Dey, T. K., & Sun, J. (2006). Defining and computing curve-skeletons with medial geodesic function. *Symposium on Geometry Processing*, *6*, 143–152.

Epstein, W., & Lovitts, B. E. (1985). Automatic and attentional components in perception of shape-at-a-slant. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(3), 355.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014–18019.

Firestone, C., & Scholl, B. J. (2014). “Please tap the shape, anywhere you like” shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological Science*, *25*(2), 377–386.

Foster, D. H. (1982). Analysis of discrete internal representations of visual pattern stimuli. *Organization and Representation in Perception. Hillsdale, NJ: Erlbaum*.

Geiger, D., Liu, T.-L., & Kohn, R. V. (2003). Representation and self-similarity of shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(1), 86–99.

Gregory, E., Landau, B., & McCloskey, M. (2011). Representation of object orientation in children: Evidence from mirror-image confusions. *Visual Cognition*, *19*(8), 1035–1062.

Gregory, E., & McCloskey, M. (2010). Mirror-image confusions: Implications for representation and processing of object orientation. *Cognition*, *116*(1), 110–129.

Haralick, R. M., & Shapiro, L. G. (1991). Glossary of computer vision terms. *Pattern Recognition*, *24*(1), 69–93.

Harris, I. M., Harris, J. A., & Caine, D. (2001). Object orientation agnosia: A failure to find the axis? *Journal of Cognitive Neuroscience*, *13*(6), 800–812.

Harrison, S. J., & Feldman, J. (2009). The influence of shape and skeletal axis structure on texture perception. *Journal of Vision*, *9*(6), 13–13.

Herbert, A. M., Humphrey, G. K., & Jolicoeur, P. (1994). The detection of bilateral symmetry: Effects of surrounding frames. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *48*(1), 140.

Hinton, G. E. (1981). The role of spatial working memory in shape perception. *Proceedings of the Third Cognitive Science Conference*, 56–60.

Hock, H. S., & Tromley, C. L. (1978). Mental rotation and perceptual uprightness. *Perception & Psychophysics*, *24*(6), 529–533.

Hoffman, D. D., & Richards, W. (1983). *Parts of recognition*.

Ittelson, W. H., Mowafy, L., & Magid, D. (1991). The perception of mirror-reflected objects. *Perception*, *20*(5), 567–584.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition*, *13*(4), 289–303.

Jolicoeur, P., & Humphrey, G. K. (1998). Perception of rotated two-dimensional and three-dimensional objects and visual shapes. *Perceptual Constancy. Why Things Look as They Do*, 69–123.

Katz, R. A., & Pizer, S. M. (2003). Untangling the Blum medial axis transform. *International Journal of Computer Vision*, *55*(2–3), 139–153.

Kimia, B. B. (2003). On the role of medial geometry in human vision. *Journal of Physiology-Paris*, *97*(2–3), 155–190.

Kovács, I., Fehér, A., & Julesz, B. (1998). Medial-point description of shape: A representation for action coding and its psychophysical correlates. *Vision Research*, *38*(15–16), 2323–2333.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105.

Large, M.-E., Mcmullen, P. A., & Hamm, J. P. (2003). The role of axes of elongation and symmetry in rotated object naming. *Perception & Psychophysics*, *65*(1), 1–19.

Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta Psychologica*, *102*(2–3), 221–245.

Lee, T.-C., Kashyap, R. L., & Chu, C.-N. (1994). Building skeleton models via 3-D medial surface axis thinning algorithms. *CVGIP: Graphical Models and Image Processing*, *56*(6), 462–478.

Leek, C. E. (1998). The analysis of orientation-dependent time costs in visual recognition. *Perception*, *27*(7), 803–816.

Leek, C. E., & Johnston, S. J. (2006). A polarity effect in misoriented object recognition: The role of polar features in the computation of orientation-invariant shape representations. *Visual Cognition*, *13*(5), 573–600.

Ling, X., & Sanocki, T. (1995). Major axes as a moderately abstract model for object recognition. *Psychological Science*, *6*(6), 370–375.

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception, & Psychophysics*, *80*(5), 1278–1289.

Maki, R. H. (1986). Naming and locating the tops of rotated pictures. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *40*(4), 368.

Marr, D. (1982). *Vison*.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *200*(1140), 269–294.

McCloskey, M. (2009). *Visual reflections: A perceptual deficit and its implications.* Oxford University Press.

McCloskey, M., Valtonen, J., & Sherman, J. C. (2006). Representing orientation: A coordinate-system hypothesis and evidence from developmental deficits. *Cognitive Neuropsychology*, *23*(5), 680–713.

McMullen, P. A., & Farah, M. J. (1991). Viewer-centered and object-centered representations in the recognition of naturalistic line drawings. *Psychological Science*, *2*(4), 275–278.

Morikawa, K. (1999). Symmetry and elongation of objects influence perceived direction of translational motion. *Perception & Psychophysics*, *61*(1), 134–143.

Ogniewicz, R. L., & Kübler, O. (1995). Hierarchic voronoi skeletons. *Pattern Recognition*, *28*(3), 343–359.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, *9*(4), 441–474.

Palmer, S. E., & Guidi, S. (2011). Mapping the perceptual structure of rectangles through goodness-of-fit ratings. *Perception*, *40*(12), 1428–1446.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 509.

Priftis, K., Rusconi, E., Umiltà, C., & Zorzi, M. (2003). Pure agnosia for mirror stimuli after right inferior parietal lesion. *Brain*, *126*(4), 908–919.

Quinlan, P. T. (1991). Differing approaches to two-dimensional shape recognition. *Psychological Bulletin*, *109*(2), 224.

Quinlan, P. T., & Humphreys, G. W. (1993). Perceptual frames of reference and two-dimensional shape recognition: Further examination of internal axes. *Perception*, *22*(11), 1343–1364.

Rezanejad, M., Downs, G., Wilder, J., Walther, D. B., Jepson, A., Dickinson, S., & Siddiqi, K. (2019). Scene categorization from contours: Medial axis based salience measures. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4116–4124.

Rezanejad, M., & Siddiqi, K. (2013). Flux graphs for 2D shape analysis. In *Shape perception in human and computer vision* (pp. 41–54). Springer.

Riddoch, M. J., & Humphreys, G. W. (1988). Description of a left/right coding deficit in a case of constructional apraxia. *Cognitive Neuropsychology*, *5*(3), 289–315.

Rock, I. (1973). *Orientation and form*. Academic Press.

Rock, I. (1983). *The logic of perception*.

Sekuler, A. B. (1996). Axis of elongation can determine reference frames for object perception. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *50*(3), 270.

Sekuler, A. B., & Swimmer, M. B. (2000). Interactions between symmetry and elongation in determining reference frames for object perception. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *54*(1), 42.

Siddiqi, K., & Pizer, S. (2008). *Medial representations: Mathematics, algorithms and applications* (Vol. 37). Springer Science & Business Media.

Singh, M., & Hoffman, D. D. (2001). Part-based representations of visual shape and implications for visual cognition. In *Advances in psychology* (Vol. 130, pp. 401–459). Elsevier.

Smith, L. B., Street, S., Jones, S. S., & James, K. H. (2014). Using the axis of elongation to align shapes: Developmental changes between 18 and 24 months of age. *Journal of Experimental Child Psychology*, *123*, 15–35.

Sutherland, N. S. (1968). Outlines of a theory of visual pattern recognition in animals and man. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *171*(1024), 297–317.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, *1*(4), 253–256.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.

Turnbull, O. H., & Mccarthy, R. A. (1996). Failure to discriminate between mirror-image objects: A case of viewpoint-independent object recognition? *Neurocase*, *2*(1), 63–72.

Van Tonder, G. J., Lyons, M. J., & Ejima, Y. (2002). Visual structure of a Japanese Zen garden. *Nature*, *419*(6905), 359–360.

Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition*, *119*(3), 325–340.

Wilder, J., Rezanejad, M., Dickinson, S., Siddiqi, K., Jepson, A., & Walther, D. B. (2019). Local contour symmetry facilitates scene categorization. *Cognition*, *182*, 307–317.

Wilson, K. D., & Farah, M. J. (2003). When does the visual system use viewpoint-invariant representations during recognition? *Cognitive Brain Research*, *16*(3), 399–415.

Winston, P. H. (1970). *Learning structural descriptions from examples*.

CURRICULUM VITA

Thitaporn (Pang) Chaisilprungraung was born on October 11th, 1990 in Bangkok, Thailand. In May 2014 she received a B.A. with honors in Cognitive Science with a minor in Linguistics at Johns Hopkins University. She subsequently began her graduate work at the Cognitive Science Department at Johns Hopkins University under the mentorship of Drs. Michael McCloskey & Soojin Park. In August 2020 she will join the Sleep Research Center at Walter Reed Army Institute of Research (WRAIR) as a Postdoctoral Research Fellow.